ARL-TR-7850 • JAN 2017

US Army Research Laboratory

Influence of Parameters of a Reactive Interatomic Potential on the Properties of Saturated Hydrocarbons

by Mark A Tschopp, Sasan Nouranian, Mike I Baskes, Steve R Gwaltney, and Mark F Horstemeyer

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-7850 ● JAN 2017

US Army Research Laboratory

Influence of Parameters of a Reactive Interatomic Potential on the Properties of Saturated Hydrocarbons

by Mark A Tschopp
Weapons and Materials Research Directorate, ARL

Sasan Nouranian
University of Mississippi, Oxford, MS

Mike I Baskes, Steve R Gwaltney, and Mark F Horstemeyer
Mississippi State University, Starkville, MS

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
January 2017	Technical Report	January 2015–June 2016

4. TITLE AND SUBTITLE
Influence of Parameters of a Reactive Interatomic Potential on the Properties of Saturated Hydrocarbons

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Mark A Tschopp, Sasan Nouranian, Mike I Baskes, Steve R Gwaltney, and Mark F Horstemeyer

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-WMM-F
Aberdeen Proving Ground, MD 21005-5069

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7850

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
primary author's email: <mark.a.tschopp.civ@mail.mil>.

14. ABSTRACT

An interatomic potential for saturated hydrocarbons using the modified embedded-atom method (MEAM), a semiempirical many-body potential based on density functional theory, was parameterized to the bond distances, bond angles, and atomization energies at 0 K of a series of alkane structures from methane to n-octane. In this work, the parameters of the MEAM potential were explored through a design of experiments and Latin hypercube sampling approach to better understand how individual MEAM parameters affected several properties of molecules (energy, bond distances, bond angles, and dihedral angles) and the relationship/correlation between various molecules in terms of these properties. The methodology can be easily extended to other potential formulations and can be useful for construction of tests and training sets for interatomic potential development.

15. SUBJECT TERMS
factorial design; Latin hypercube; hydrocarbon; interatomic potential; molecular dynamics; optimization; carbon; hydrogen

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
Unclassified	Unclassified	Unclassified	UU	62	Mark A Tschopp 410-306-0855

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures iv

List of Tables vii

1. Introduction 1

2. Simulation Methodology 3

2.1 Modified Embedded-Atom Method Theory 3

2.1.1 Embedding Energy Function 3

2.1.2 Screening Factor 8

2.1.3 Modified Embedded-Atom Method Parameterization 9

2.2 Fractional Factorial Design 9

2.3 Tests: Molecules and Property 14

3. Simulation Results 16

3.1 Analysis of Variance 16

3.2 Interatomic Potential Parameter Sensitivity via Analysis of Variance 18

3.3 Property Similarity for Different Molecules/Potentials 33

4. Summary and Conclusion 44

5. References 46

List of Symbols, Abbreviations, and Acronyms 50

Distribution List 51

List of Figures

Fig. 1 Example of some of the molecules used in the parameterization of the C–H modified embedded-atom method (MEAM) potential and in the property assessment in this study 15

Fig. 2 Example of some of the molecule configuration paths used for the energy-distance relationships (as distance is increased in horizontal direction), which was then used in the parameterization of the C–H MEAM potential and in property assessment in this study 16

Fig. 3 Summary of ANOVA results for the energies of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. 20

Fig. 4 Summary of ANOVA results for the C–H bond distances of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond distance for all C–H bonds was used for those molecules with multiple C–H bonds. 21

Fig. 5 Summary of ANOVA results for the C–C bond distances of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond distance for all C–C bonds was used for those molecules with multiple C–C bonds. 22

Fig. 6 Summary of ANOVA results for the C–C–C bond angles of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond angle for all C–C–C bond angles was used for those molecules with multiple C–C–C bond angles. 24

Fig. 7 Summary of ANOVA results for the C–C–H bond angles of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond angle for all C–C–H bond angles was used for those molecules with multiple C–C–H bond angles. 25

Fig. 8 Summary of ANOVA results for the H–C–H bond angles of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond angle for all H–C–H bond angles was used for those molecules with multiple H–C–H bond angles.26

Fig. 9 Summary of ANOVA results for the C–C–C–C dihedral angles of all molecules examined. Since there are primarily 2 dihedral angles in the tested molecules (60° and 180°), the absolute value of the minimum difference between each dihedral angle and these 2 reference angles was used. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean dihedral angle for all C–C–C–C angles was used for those molecules with multiple C–C–C–C dihedral angles. …… 27

Fig. 10 Summary of ANOVA results for the C–C–C–H dihedral angles of all molecules examined. Since there are primarily 2 dihedral angles in the tested molecules (60° and 180°), the absolute value of the minimum difference between each dihedral angle and these 2 reference angles was used. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean dihedral angle for all C–C–C–H angles was used for those molecules with multiple C–C–C–H dihedral angles. …… 28

Fig. 11 Summary of ANOVA results for the H–C–C–H dihedral angles of all molecules examined. Since there are primarily 2 dihedral angles in the tested molecules (60° and 180°), the absolute value of the minimum difference between each dihedral angle and these 2 reference angles was used. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean dihedral angle for all H–C–C–H angles was used for those molecules with multiple H–C–C–H dihedral angles. …… 29

Fig. 12 Summary of ANOVA results for the energy vs. distance relationships for various molecules and paths. In this contour plot, the mean absolute error between the calculated MEAM curve and first principles data at the same distances was used as a response variable. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. …… 31

Fig. 13 Summary of ANOVA results for the energy vs. distance relationships for various molecules and paths. In this contour plot, the root mean square error between the calculated MEAM curve and first principles data at the same distances was used as a response variable. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. 32

Fig. 14 Correlation coefficient R for the energies of all molecules examined (for the Latin hypercube sampling design). In this plot, each response is plotted against all of the other responses. A high correlation coefficient indicates the strength of the linear dependence between the 2 responses (i.e., a high correlation, R=1, for responses plotted against themselves). 35

Fig. 15 Correlation coefficient R for the energies of all molecules examined (for the fractional factorial design case). In this plot, each response is plotted against all of the other responses. A high correlation coefficient indicates the strength of the linear dependence between the 2 responses (i.e., a high correlation, R=1, for responses plotted against themselves). 36

Fig. 16 Correlation coefficient R for the C–H bond distances of all molecules examined. The mean bond distance for all C–H bonds was used for those molecules with multiple C–H bonds. 38

Fig. 17 Correlation coefficient R for the C–C bond distances of all molecules examined. The mean bond distance for all C–C bonds was used for those molecules with multiple C–C bonds. 39

Fig. 18 Correlation coefficient R for the C–C–C bond angles of all molecules examined. The mean bond angle for all C–C–C bond angles was used for those molecules with multiple C–C–C bond angles. 41

Fig. 19 Correlation coefficient R for the C–C–H bond angles of all molecules examined. The mean bond angle for all C–C–H bond angles was used for those molecules with multiple C–C–H bond angles. 42

Fig. 20 Correlation coefficient R for the H–C–H bond angles of all molecules examined. The mean bond angle for all H–C–H bond angles was used for those molecules with multiple H–C–H bond angles. 43

List of Tables

Table 1 Example of fractional factorial designs (2^{5-2} and 2^{5-1} designs) vs. a full factorial design (2^5 design) for 5 factors: a, b, c, d, e 11

Table 2 Resolution IV fractional factorial design for 26 factors..... 12

Table 3 The 26 MEAM parameters and their assigned values in the 2^{26-20} fractional factorial design 13

Table 4 ANOVA analysis of energy of CH_4 molecule..... 18

INTENTIONALLY LEFT BLANK.

1. Introduction

The embedded-atom method (EAM), developed by Daw and Baskes[1,2] in the early 1980s, is a semiempirical N-body potential useful for the atomistic simulations of metal systems. It has successfully been used to calculate the energetics and structures of complex metallic systems involving free surfaces, defects, grain boundaries, etc.[3] The potential was later modified by Baskes[4,5] to include the directionality of bonding in covalent materials such as silicon and germanium,[6] leading to the MEAM[5] introduced in 1992. It has undergone several modifications and enhancements since then to include, for example, second nearest-neighbor (2NN) interactions[7–9] and, more recently, a multistate formalism.[10] The unique feature of the MEAM formalism is its ability to reproduce the physical properties of a large number of face-centered cubic (FCC),[9,11] body-centered cubic (BCC),[8,12] hexagonal close-packed (HCP),[13,14] and diamond cubic[15] crystal structures in unary, binary, ternary, and higher order[16] metal systems with the same semiempirical formalism. MEAM is also both reliable and transferable[17] in the sense that it accurately reproduces the physical properties of the element or alloy (reliability) and performs reasonably well under circumstances other than the ones used for its parameterization (transferability).[17]

The MEAM formalism has traditionally been used for pure metals and impurities, binary and ternary alloys, and hydride, carbide, and nitride metal systems with great success,[18] as outlined in a review article by Horstemeyer.[19] In addition, complex nanostructured systems have been studied using various MEAM-based potentials. For example, Xiao et al.[20] calculated the interaction of carbon nanotubes with Ni nanoparticles, and Uddin et al.[21] recently studied the mechanical properties of carbon nanotube-Ni composites using the MEAM potential. The MEAM formalism has been extended to saturated hydrocarbons with the ultimate aim of capturing the energetics and geometries of commercially important hydrocarbon-based polymers (polyolefins) such as polyethylene and polypropylene. Potentials such as MM3,[22–24] MM4,[25] DREIDING,[26] first-[27] and second-generation reactive empirical bond order (REBO),[28] reactive force field (ReaxFF),[29] charge-optimized many-body (COMB) potential,[30,31] and condensed-phase optimized molecular potentials for atomistic simulation studies (COMPASS)[32] have been used for hydrocarbon simulations, but of these potentials, only REBO, ReaxFF, and COMB are reactive and can allow for bond breaking. Furthermore, most of these potentials are

not suitable for hydrocarbon-metal systems, with only ReaxFF[33–36] and COMB[37] having been used in the past to study hydrocarbon-metal interactions. Liang et al.[38] reviewed the use of reactive potentials for advanced atomistic simulations.

A new set of parameters within the MEAM framework was recently developed to describe the interactions and equilibrium geometries of saturated hydrocarbons, specifically bond distances, bond angles, and atomization energies at 0 K. It was shown that MEAM gives a comparable or more accurate reproduction of these properties relative to experimental and first-principles data in comparison with REBO and ReaxFF. It was also able to reproduce the potential energy curves of H_2, CH, and C_2 diatomics and $(H_2)_2$, $(CH_4)_2$, $(C_2H_6)_2$, and $(C_3H_8)_2$ dimer configurations and predict the pressure-volume-temperature (PVT) relationships of a series of select methane, ethane, propane, and butane systems in a reasonable agreement with the experimental data. The energetics of C–H and C–C bond breaking in methane and ethane and the heat of reaction for select chemical reactions have also been explored, and it was found that MEAM gives reasonable predictions of the energies associated with these 2 major chemical reactions in saturated hydrocarbons.

The present work is associated with the initial development of the C–H MEAM potential.[39] In parameterizing this potential, a number of other studies were performed to understand the relative influence of various parameters on the training set of properties for these different molecules and configurations. Moreover, further analysis can shed light on the relationship between different properties and the similarity/dissimilarity between different molecules within the training set. In doing these analyses, reduced sampling strategies were investigated, including fractional factorial designs and Latin hypercube sampling designs. A few of these studies and their analyses are presented herein as methods and tools for both understanding and developing interatomic potentials. This work builds upon prior studies by the various authors.[39–41] This report is organized in the following manner. In Section 2 the theory of the MEAM formalism is reviewed. In Section 3, the simulation results for the 2 studies is presented and analyzed. Last, the summary follows in Section 4.

2. Simulation Methodology

2.1 Modified Embedded-Atom Method Theory

In the EAM and MEAM formalisms[1,2,5] the total energy of a system of atoms (E_{tot}) is given by

$$E_{tot} = \sum_i \left[F_{\tau_i}\left(\bar{\rho}_i\right) + \frac{1}{2} \sum_{j(\neq i)} S_{ij} \phi_{\tau_i \tau_j}\left(R_{ij}\right) \right], \tag{1}$$

where F_{τ_i} is the embedding energy function for element type τ_i, which is defined as the energy required to embed an atom of element type τ_i in the background electron density $\bar{\rho}_i$ at site i, S_{ij} is the screening factor between atoms at sites i and j (defined in Eqs. 25 and 28), and $\phi_{\tau_i \tau_j}$ is the pair interaction between atoms of element types τ_i and τ_j at sites i and j at the separation distance of R_{ij}. To emphasize the multi-component nature of the model, the element type of the atom at site i is denoted as τ_i in this manuscript to distinguish it from site designation i, and the screening factor is explicitly separated from the pair potential. The following subsubsections describe the calculation of the embedding energy function and screening factor for the MEAM formalism.

2.1.1 Embedding Energy Function

The embedding function is given by the specific simple form

$$F_{\tau_i}\left(\bar{\rho}_i\right) = \begin{cases} A_{\tau_i} E^0_{\tau_i} \dfrac{\bar{\rho}_i}{\bar{\rho}^0_{\tau_i}} \left(\ln \dfrac{\bar{\rho}_i}{\bar{\rho}^0_{\tau_i}} \right) & \text{if } \bar{\rho}_i \geq 0 \\ -A_{\tau_i} E^0_{\tau_i} \dfrac{\bar{\rho}_i}{\bar{\rho}^0_{\tau_i}} & \text{if } \bar{\rho}_i < 0 \end{cases}, \tag{2}$$

where A_{τ_i} is a scaling factor, $E^0_{\tau_i}$ is the sublimation (cohesive) energy, and $\bar{\rho}^0_{\tau_i}$ is the background electron density for the reference structure of the atom of element type τ_i at site i. For most elements, the reference structure is the equilibrium structure of the element in its reference state. However, the reference structure of carbon is taken as diamond. We will denote the properties of the equilibrium reference state with a superscript zero. The analytic continuation of the embedding function for negative electron densities was considered as a computational convenience to pre-

vent systems from entering this unphysical regime. The origin of negative electron densities arises in Eq. 8. The MEAM formalism introduces directionality in bonding between atoms through decomposing $\bar{\rho}_i$ into spherically symmetric ($\rho_i^{(0)}$) and angular ($\rho_i^{(1)}$, $\rho_i^{(2)}$, and $\rho_i^{(3)}$) partial electron densities[5,18,42] as given by

$$\rho_i^{(0)} = \sum_{j \neq i} S_{ij} \rho_{\tau_i}^{a(0)} \left(R_{ij} \right), \tag{3}$$

$$\left(\rho_i^{(1)} \right)^2 = \frac{\sum\limits_{\alpha} \left[\sum\limits_{j \neq i} \frac{R_{ij}^{\alpha}}{R_{ij}} S_{ij} t_{\tau_j}^{(1)} \rho_{\tau_j}^{a(1)} \left(R_{ij} \right) \right]^2 \rho_i^{(0)}}{\sum\limits_{j \neq i} S_{ij} \left(t_{\tau_j}^{(1)} \right)^2 \rho_{\tau_j}^{a(0)} \left(R_{ij} \right)}, \tag{4}$$

$$\begin{aligned} \left(\rho_i^{(2)} \right)^2 = & \frac{\left\{ \sum\limits_{\alpha,\beta} \left[\sum\limits_{j \neq i} \frac{R_{ij}^{\alpha} R_{ij}^{\beta}}{R_{ij}^2} S_{ij} t_{\tau_j}^{(2)} \rho_{\tau_j}^{a(2)} \left(R_{ij} \right) \right]^2 \right\} \rho_i^{(0)}}{\sum\limits_{j \neq i} S_{ij} \left(t_{\tau_j}^{(2)} \right)^2 \rho_{\tau_j}^{a(0)} \left(R_{ij} \right)} + \\ & \frac{-\frac{1}{3} \left\{ \left[\sum\limits_{j \neq i} S_{ij} t_{\tau_j}^{(2)} \rho_{\tau_j}^{a(2)} \left(R_{ij} \right) \right]^2 \right\} \rho_i^{(0)}}{\sum\limits_{j \neq i} S_{ij} \left(t_{\tau_j}^{(2)} \right)^2 \rho_{\tau_j}^{a(0)} \left(R_{ij} \right)}, \end{aligned} \tag{5}$$

$$\begin{aligned} \left(\rho_i^{(3)} \right)^2 = & \frac{\left\{ \sum\limits_{\alpha,\beta,\gamma} \left[\sum\limits_{j \neq i} \frac{R_{ij}^{\alpha} R_{ij}^{\beta} R_{ij}^{\gamma}}{R_{ij}^3} S_{ij} t_{\tau_j}^{(3)} \rho_{\tau_j}^{a(3)} \left(R_{ij} \right) \right]^2 \right\} \rho_i^{(0)}}{\sum\limits_{j \neq i} S_{ij} \left(t_{\tau_j}^{(3)} \right)^2 \rho_{\tau_j}^{a(0)} \left(R_{ij} \right)} + \\ & \frac{-\frac{3}{5} \left\{ \sum\limits_{\alpha} \left[\sum\limits_{j \neq i} \frac{R_{ij}^{\alpha}}{R_{ij}} S_{ij} t_{\tau_j}^{(3)} \rho_{\tau_j}^{a(3)} \left(R_{ij} \right) \right]^2 \right\} \rho_i^{(0)}}{\sum\limits_{j \neq i} S_{ij} \left(t_{\tau_j}^{(3)} \right)^2 \rho_{\tau_j}^{a(0)} \left(R_{ij} \right)}. \end{aligned} \tag{6}$$

$\rho_{\tau_j}^{a(h)}$ ($h = \{0, 1, 2, 3\}$) indicate the atomic electron densities from atom of element type τ_j at site j at distance R_{ij} from site i. R_{ij}^{α}, R_{ij}^{β}, and R_{ij}^{γ} represent the α, β, and γ components of the distance vector between atoms at sites i and j, respectively, and $t_{\tau_j}^{(h)}$ ($h = \{1, 2, 3\}$) are adjustable element-dependent parameters. The equivalence between these expressions and an expansion in Legendre polynomials has been discussed previously.[5] As previously discussed, we have carefully denoted the element types of the atoms, and separated the screening from the atomic electron densities. Equation 3 is the simple linear superposition of atomic densities of the EAM formalism,[1,2] and Eqs. 4–6 reduce to more familiar forms in the original MEAM paper by Baskes[5] for a single-component system. The partial electron densities can be combined in different ways to give the total background electron density at site i ($\bar{\rho}_i$). Here, we adopt one of the most widely used forms,[16,17,43] which is given by

$$\bar{\rho}_i = \rho_i^{(0)} G\left(\Gamma_i\right), \tag{7}$$

$$G\left(\Gamma_i\right) = \begin{cases} \sqrt{1+\Gamma_i} & \text{if } \Gamma_i \geq -1 \\ -\sqrt{|1+\Gamma_i|} & \text{if } \Gamma_i < -1 \end{cases}, \tag{8}$$

$$\Gamma_i = \sum_{h=1}^{3} \bar{t}_i^{(h)} \left[\frac{\rho_i^{(h)}}{\rho_i^{(0)}}\right]^2, \tag{9}$$

$$\bar{t}_i^{(h)} = \frac{1}{\rho_i^{(0)}} \sum_{j \neq i} t_{\tau_j}^{(h)} \rho_{\tau_j}^{a(0)} S_{ij}. \tag{10}$$

In the absence of angular contributions to the density, $\Gamma_i = 0$, $G(\Gamma_i) = 1$, the model reduces to the EAM formalism. For systems with negative $t_{\tau_j}^{(h)}$ values in certain geometries, $\Gamma_i < -1$, and for computational convenience, we perform an analytic continuation of $G\left(\Gamma_i\right)$. We choose to do this by allowing $G\left(\Gamma_i\right)$ and, hence, $\bar{\rho}_i$ to become less than zero.

If we apply Eqs. 7 and 9 to the equilibrium reference structure, we obtain

$$\bar{\rho}_\tau^0 = Z_\tau^0 \rho_\tau^0 G\left(\Gamma_\tau^0\right), \tag{11}$$

$$\Gamma_\tau^0 = \sum_{h=1}^{3} t_\tau^{(h)} s_\tau^{(h)} \left(\frac{1}{Z_\tau^0}\right)^2, \tag{12}$$

where we have assumed that the reference structure has only first nearest-neighbor (1NN) interactions. In Eq. 11, ρ_τ^0 is an element-dependent electron density scaling factor, and Z_τ^0 is the 1NN coordination number of the reference structure. $s_\tau^{(h)}$ ($h = \{1, 2, 3\}$) are "shape factors" that depend on the reference structure for element type τ. The shape factors are given in the original MEAM paper by Baskes.[5] The atomic electron density for element type τ is calculated from

$$\rho_\tau^{a(h)}(R) = \rho_\tau^0 e^{-\beta_\tau^{(h)}\left(\frac{R}{R_\tau^0}-1\right)}, \tag{13}$$

where $\beta_\tau^{(h)}$ ($h = \{0, 1, 2, 3\}$) are adjustable element-dependent parameters, and R_τ^0 is the nearest-neighbor distance in the equilibrium reference structure for the element type τ.

The pair interaction for like atoms of element type τ can be calculated using a 1NN[5] or 2NN[7,18] formalism. In this work, the former is used[16,17] and is given by

$$\phi_{\tau\tau}(R) = \frac{2}{Z_\tau^0}\left\{E_\tau^u(R) - F_\tau\left[\bar{\rho}_\tau^{ref}(R)\right]\right\}. \tag{14}$$

In this equation $\bar{\rho}_\tau^{ref}(R)$ is the background electron density in the reference structure evaluated from Eqs. 7–10 at a nearest-neighbor distance of R and is given by

$$\bar{\rho}_\tau^{ref}(R) = Z_\tau^0 \rho_\tau^0 G\left(\Gamma_\tau^{ref}\right), \tag{15}$$

$$\Gamma_\tau^{ref} = \sum_{h=1}^{3} t_\tau^{(h)} s_\tau^{(h)} \left(\frac{\rho_\tau^{a(h)}}{Z_\tau^0 \rho_\tau^{a(0)}}\right)^2, \tag{16}$$

and E_τ^u is the universal equation of state (UEOS) of Rose et al.[44] for element type τ given by

$$E_\tau^u(R) = -E_\tau^0 \left[1 + a^* + \delta \frac{R_\tau^0}{R} (a^*)^3\right] e^{-a^*}, \tag{17}$$

$$a^* = \alpha_\tau^0 \left(\frac{R}{R_\tau^0} - 1\right), \tag{18}$$

$$\delta = \begin{cases} \delta_\tau^a & \text{if } a^* \geq 0 \\ \delta_\tau^r & \text{if } a^* < 0 \end{cases}, \tag{19}$$

$$\alpha_\tau^0 = \sqrt{\frac{9K_\tau^0 \Omega_\tau^0}{E_\tau^0}}, \tag{20}$$

or

$$\alpha_\tau^0 = \sqrt{\frac{k_\tau^0}{E_\tau^0}} R_\tau^0. \tag{21}$$

In Eqs. 17–21, K_τ^0, k_τ^0, and Ω_τ^0 are the bulk modulus, spring constant, and the atomic volume of the reference structure, respectively, and δ is an adjustable, element-dependent parameter that has 2 components, attractive δ_τ^a and repulsive δ_τ^r. Equation 20 is used when the reference structure is a 3-dimensional crystal, and Eq. 21 is used when the reference structure is a diatomic.

The pair interaction for unlike atoms of element types τ and υ is similarly obtained from the reference structure of the unlike atoms. For this work, the reference structure is taken as the heteronuclear diatomic, which gives

$$\phi_{\tau\upsilon}(R) = \frac{1}{Z_{\tau\upsilon}^0} \left\{2E_{\tau\upsilon}^u(R) - F_\tau\left[\bar{\rho}_\upsilon^d(R)\right] - F_\upsilon\left[\bar{\rho}_\tau^d(R)\right]\right\}, \tag{22}$$

where $Z_{\tau\upsilon}^0 = 1$ is the coordination number for the diatomic and

$$\bar{\rho}_{\tau}^{d}\left(R\right)=\rho_{\tau}^{a(0)}G\left(\Gamma_{\tau}^{d}\right), \tag{23}$$

$$\Gamma_{\tau}^{d}=\sum_{h=1}^{3}t_{\tau}^{(h)}s_{d}^{(h)}\left(\frac{\rho_{\tau}^{a(h)}}{\rho_{\tau}^{a(0)}}\right)^{2}, \tag{24}$$

where the shape factors $s_d^{(h)}$ are those for a diatomic. The UEOS $E_{\tau\upsilon}^{u}$ is given by Eqs. 17–21 using parameters $E_{\tau\upsilon}^{0}$, $R_{\tau\upsilon}^{0}$, $k_{\tau\upsilon}^{0}$, $\delta_{\tau\upsilon}^{a}$, and $\delta_{\tau\upsilon}^{r}$.

2.1.2 Screening Factor

The screening factor S_{ij} is defined as the product of all screening factors S_{ikj}, where the interaction between atoms at sites i and j are screened by neighboring atoms at site k as given by

$$S_{ij}=\prod_{k\neq i,j}S_{ikj}. \tag{25}$$

If it is assumed that all 3 sites i, j, and k lie on an ellipse on the xy-plane with sites i and j on the x-axis, the following equation can be derived:

$$x^{2}+\frac{1}{C}y^{2}=\left(\frac{1}{2}R_{ij}\right)^{2}, \tag{26}$$

where

$$C_{ikj}=\frac{2\left(X_{ik}+X_{kj}\right)-\left(X_{ik}-X_{kj}\right)^{2}-1}{1-\left(X_{ik}-X_{kj}\right)^{2}}. \tag{27}$$

In Eq. 27, $X_{ik}=\left(R_{ik}/R_{ij}\right)^{2}$ and $X_{kj}=\left(R_{kj}/R_{ij}\right)^{2}$. The screening factor S_{ikj} for like atoms is defined as

$$S_{ikj}=f_{c}\left(\frac{C_{ikj}-C_{min}\left(\tau_{i},\tau_{k},\tau_{j}\right)}{C_{max}\left(\tau_{i},\tau_{k},\tau_{j}\right)-C_{min}\left(\tau_{i},\tau_{k},\tau_{j}\right)}\right), \tag{28}$$

where $C_{min}(\tau_i, \tau_k, \tau_j)$ and $C_{max}(\tau_i, \tau_k, \tau_j)$ determine the extent of screening of atoms of element type τ at sites i and j by an atom at site k. Similar expressions can be written for the screening of unlike atoms. The smooth cutoff function f_c is defined as

$$f_c(x) = \begin{cases} 1 & \text{if } x \geq 1 \\ \left[1 - (1-x)^4\right]^2 & \text{if } 0 < x < 1 \\ 0 & \text{if } x \leq 0 \end{cases} . \tag{29}$$

$S_{ij} = 1$ means that the interaction between atoms at sites i and j is not screened, while $S_{ij} = 0$ means the interaction is completely screened.

2.1.3 Modified Embedded-Atom Method Parameterization

The MEAM formalism presented in Eqs. 1–29 requires 16 independent model parameters for each element type τ (i.e., E_τ^0, R_τ^0, α_τ^0, δ_τ^a, and δ_τ^r for the UEOS [Eq. 17]; $\beta_\tau^{(0)}$, $\beta_\tau^{(1)}$, $\beta_\tau^{(2)}$, $\beta_\tau^{(3)}$, $t_\tau^{(1)}$, $t_\tau^{(2)}$, $t_\tau^{(3)}$, and ρ_τ^0 for the electron densities [Eqs. 3–13]; A_τ for the embedding function F_τ [Eq. 2]; and C_{min} and C_{max} for the screening factor [Eqs. 25–29]). In the current MEAM formalism for a single element, the model is independent of ρ_τ^0; hence, $\rho_\tau^0 = 1$ is taken for one of the elements. For a diatomic composed of elements τ and υ, 13 additional independent parameters are required (i.e., $E_{\tau\upsilon}^0$, $R_{\tau\upsilon}^0$, $\alpha_{\tau\upsilon}^0$, $\delta_{\tau\upsilon}^a$, and $\delta_{\tau\upsilon}^r$, 4 C_{min}, and 4 C_{max} values). A small molecule database was used in the fitting of the C–H MEAM interatomic potential. Further details of the parameterization are provided in Nouranian et al.[39] The simulations were run using the molecular dynamics code, DYNAMO.[45]

2.2 Fractional Factorial Design

A design of experiments approach was used to assess the influence of each parameter on the small molecule database that was used for tuning the C–H MEAM interatomic potential. A fractional factorial design was chosen to identify the significant main effects on a large number of responses: molecular atomization energies, average bond lengths (C–C, C–H, H–H), average bond angles (C–C–C, C–C–H, and H–C–H), average dihedral angles (C–C–C–C, C–C–C–H, H–C–C–H), and difference in the energy vs. distance relationship from first principles calculations for several molecules (2C, 2H, C–H, CH_4, and C_2H_6) and paths. Table 1 shows an example of a fractional factorial experiment for 5 factors (in this study, we used 26

parameters) with 2 levels (-1 and $+1$). The 2^5 factorial is a full factorial design and requires 32 runs to test all combinations. For fractional factorial designs, only a fraction of these runs are used, which results in certain effects being confounded with each other (e.g., it may not be possible to understand whether a response is due to factor a or due to an interaction between factor b and factor c). The resolution of a fractional factorial design details the extent of confounding within the design. In Table 1, the 2^{5-1} fractional factorial design is a Resolution IV design, where the main effects are not confounded with any other main effects or 2-factor interactions, but the 2 factor interactions are confounded with each other. The 2^{5-2} fractional factorial design is a Resolution III design that only requires 8 runs (one-quarter of a full factorial design) but where the main effects may be confounded with 2-factor interactions. In the present study, a 2^{26-20} fractional factorial design is used (Table 2), which requires 64 runs. This factorial design is a Resolution IV design, so the main effects are unconfounded with each other or with 2-factor interactions, providing information about which parameters significantly affect the responses being examined. Table 3 shows the 26 MEAM parameters used in the present study and the -1 and $+1$ levels for each factor.

Table 1 Example of fractional factorial designs (2^{5-2} and 2^{5-1} designs) vs. a full factorial design (2^5 design) for 5 factors: a, b, c, d, e

2^{5-2} Factorial Design						2^{5-1} Factorial Design						2^5 Factorial Design					
Run	a	b	c	d	e	Run	a	b	c	d	e	Run	a	b	c	d	e
1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1
2	-1	-1	1	1	-1	2	-1	-1	-1	1	-1	2	-1	-1	-1	-1	1
3	-1	1	-1	1	-1	3	-1	-1	1	-1	-1	3	-1	-1	-1	1	-1
4	-1	1	1	-1	1	4	-1	-1	1	1	1	4	-1	-1	-1	1	1
5	1	-1	-1	1	1	5	-1	1	-1	-1	-1	5	-1	-1	1	-1	-1
6	1	-1	1	-1	-1	6	-1	1	-1	1	1	6	-1	-1	1	-1	1
7	1	1	-1	-1	-1	7	-1	1	1	-1	1	7	-1	-1	1	1	-1
8	1	1	1	1	1	8	-1	1	1	1	-1	8	-1	-1	1	1	1
						9	1	-1	-1	-1	-1	9	-1	1	-1	-1	-1
						10	1	-1	-1	1	1	10	-1	1	-1	-1	1
						11	1	-1	1	-1	1	11	-1	1	-1	1	-1
						12	1	-1	1	1	-1	12	-1	1	-1	1	1
						13	1	1	-1	-1	1	13	-1	1	1	-1	-1
						14	1	1	-1	1	-1	14	-1	1	1	-1	1
						15	1	1	1	-1	-1	15	-1	1	1	1	-1
						16	1	1	1	1	1	16	-1	1	1	1	1
												17	1	-1	-1	-1	-1
												18	1	-1	-1	-1	1
												19	1	-1	-1	1	-1
												20	1	-1	-1	1	1
												21	1	-1	1	-1	-1
												22	1	-1	1	-1	1
												23	1	-1	1	1	-1
												24	1	-1	1	1	1
												25	1	1	-1	-1	-1
												26	1	1	-1	-1	1
												27	1	1	-1	1	-1
												28	1	1	-1	1	1
												29	1	1	1	-1	-1
												30	1	1	1	-1	1
												31	1	1	1	1	-1
												32	1	1	1	1	1

Table 2 Resolution IV fractional factorial design for 26 factors

Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1
3	-1	-1	-1	-1	1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1
4	-1	-1	-1	-1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1
5	-1	-1	-1	1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	1	1	1	1	-1
6	-1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	1	-1
7	-1	-1	-1	1	1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1
8	-1	-1	-1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	1
9	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1	-1	-1	1
10	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	1
11	-1	-1	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1
12	-1	-1	1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	-1
13	-1	-1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	1	1	1
14	-1	-1	1	1	-1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
15	-1	-1	1	1	1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	1	1	-1	-1	-1
16	-1	-1	1	1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1
17	-1	1	-1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1
18	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
19	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1
20	-1	1	-1	-1	1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1
21	-1	1	-1	1	-1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	1	-1	1	1
22	-1	1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
23	-1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	1	-1	1	-1	-1
24	-1	1	-1	1	1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1
25	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1	-1	-1
26	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1
27	-1	1	1	-1	1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	1	-1	-1	1	1
28	-1	1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
29	-1	1	1	1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	1	-1	-1	1	-1
30	-1	1	1	1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1
31	-1	1	1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	1	1	-1	1
32	-1	1	1	1	1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	-1	1	1	-1	1
33	1	-1	-1	-1	-1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	1	-1	-1	1	-1
34	1	-1	-1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1
35	1	-1	-1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	1
36	1	-1	-1	-1	1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	-1	1	1	-1	1
37	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	-1	1	1	-1	-1
38	1	-1	-1	1	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1
39	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	1	1
40	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
41	1	-1	1	-1	-1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	1
42	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
43	1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	1	-1	-1
44	1	-1	1	-1	1	1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1
45	1	-1	1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1
46	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
47	1	-1	1	1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1
48	1	-1	1	1	1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1
49	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	-1	1	1	1
50	1	1	-1	-1	-1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
51	1	1	-1	-1	1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	1	1	-1	-1	-1
52	1	1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1
53	1	1	-1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	1	1	-1	-1	1
54	1	1	-1	1	-1	1	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	1
55	1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1
56	1	1	-1	1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	-1
57	1	1	1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	1	1	1	1	-1
58	1	1	1	-1	-1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	-1
59	1	1	1	-1	1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1
60	1	1	1	-1	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	1
61	1	1	1	1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
62	1	1	1	1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1
63	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 3 The 26 MEAM parameters and their assigned values in the 2^{26-20} fractional factorial design

x_i	Parameter		-1	+1
x_1	α_C^0	(alpha1)	3.500	3.700
x_2	$\beta_C^{(0)}$	(b01)	4.190	4.210
x_3	$\beta_C^{(1)}$	(b11)	4.490	4.510
x_4	$\beta_C^{(2)}$	(b21)	4.290	4.310
x_5	$\beta_C^{(3)}$	(b31)	4.170	4.190
x_6	R_C^0	(alat1)	3.315	3.425
x_7	E_C^0	(esub1)	7.170	7.570
x_8	A_C	(asub1)	0.540	0.740
x_9	α_H^0	(alpha2)	2.029	2.049
x_{10}	$\beta_H^{(0)}$	(b02)	2.710	2.730
x_{11}	$\beta_H^{(1)}$	(b12)	2.035	2.055
x_{12}	$\beta_H^{(2)}$	(b22)	2.240	2.260
x_{13}	$\beta_H^{(3)}$	(b32)	2.990	3.010
x_{14}	R_H^0	(alat2)	0.730	0.750
x_{15}	E_H^0	(esub2)	2.163	2.563
x_{16}	A_H	(asub2)	2.400	2.600
x_{17}	ρ_H^0	(rozero2)	1.750	1.850
x_{18}	E_{CH}^0	(deltas12)	1.970	2.270
x_{19}	R_{CH}^0	(res12)	1.010	1.030
x_{20}	α_{CH}^0	(alphas12)	3.190	3.210
x_{21}	δ_C^a	(attrac11)	-0.01	0.01
x_{22}	δ_C^r	(repuls11)	0.00	0.01
x_{23}	δ_{CH}^a	(attrac12)	0.04	0.06
x_{24}	δ_{CH}^r	(repuls12)	0.04	0.06
x_{25}	δ_H^a	(attrac22)	-0.01	0.01
x_{26}	δ_H^r	(repuls22)	0.04	0.06

The next step is to assign the interatomic potential parameters to the 26 factors and to select appropriate numeric values for the potential parameters (i.e., the -1 and $+1$ factor levels). In the following study, the MEAM parameters identified in prior work[39] were used as baseline values and the -1 and $+1$ factor levels were chosen to bracket these values.

2.3 Tests: Molecules and Property

The last step is to define the responses used for the present study. Figure 1 shows an example of some of the molecules used for parameterization of the potential as well as used herein. The molecule set represents a diversity of potential structures for the C–H MEAM potential: dimers, alkanes, alkenes, alkynes, cycloalkanes, aromatics, and radicals, some of which are depicted in Fig. 1. The various properties explored for each molecule include 0 K energy, C–C and C–H bond distances, C–C–C, C–C–H, and H–C–H bond angles, and C–C–C–C, C–C–C–H, and H–C–C–H dihedral angles. For the case of molecules with multiple values for bond distances, bond angles, and dihedral angles, the average value for all bond/angle/dihedral types of that nature are used. Another property that is quantified within is the energy vs. distance relationship for a few dimers and small molecules. For example, Fig. 2 depicts 2 C_2H_6 molecules that are oriented in different ways with respect to each other. Each of these molecules is then separated by different distances while constraining the atoms and the energy is calculated. This type of calculation is performed for H_2, CH_4, and C_2H_6, and was compared to first principles results as discussed in Nouranian et al.[39]

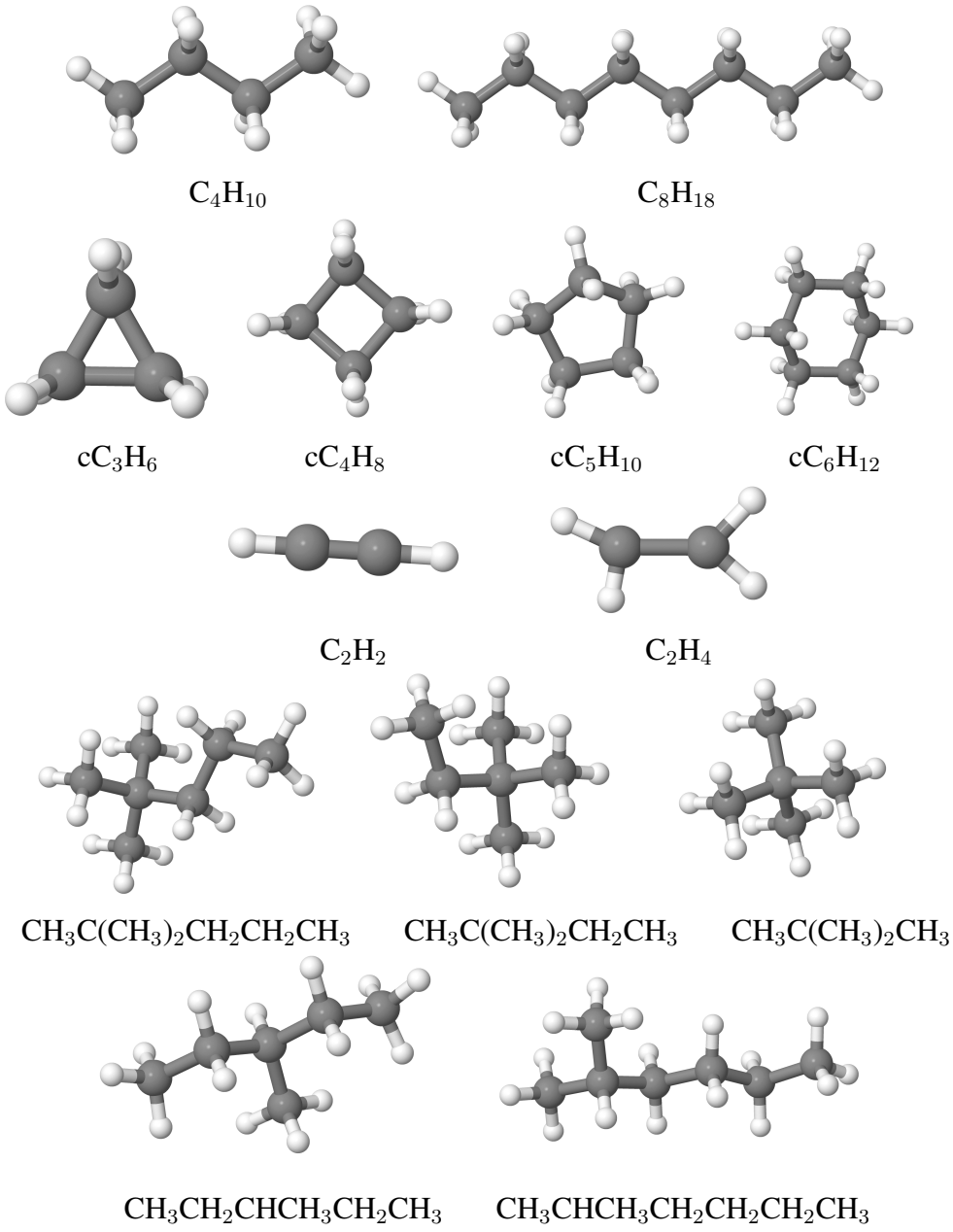

Fig. 1 Example of some of the molecules used in the parameterization of the C–H MEAM potential and in the property assessment in this study

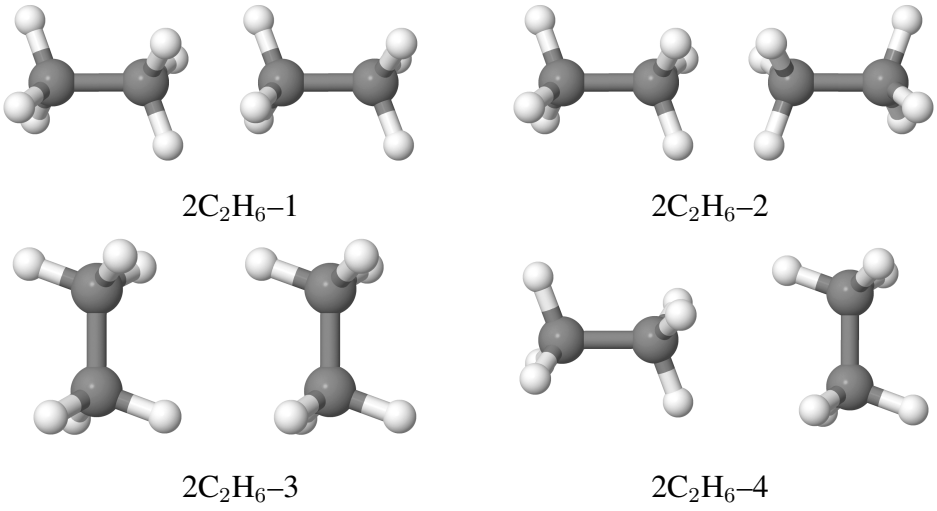

Fig. 2 Example of some of the molecule configuration paths used for the energy-distance relationships (as distance is increased in horizontal direction), which was then used in the parameterization of the C–H MEAM potential and in property assessment in this study

3. Simulation Results

3.1 Analysis of Variance

Factorial or fraction factorial designs can be analyzed by describing the response variables in terms of an overall mean, a contribution from the treatment (i.e., due to the different factors and factor levels), and a contribution from random error—in the case of minimum energy molecules herein, the responses for each set of parameters is unique and so only one simulation is required. The variability in the responses is a function of the treatment and, hence, the various factors and factor levels. So, an effective method for analyzing factorial experiments is to measure the total deviation from the mean response (i.e., sum of squares) and partition this into the different factors (i.e., the analysis of variance [ANOVA] method). This partitioning of the variability into the different factors enables testing the hypothesis of whether any particular factor has a significant influence on the total variability; this is often expressed through an F ratio test statistic and a p-value.

The ANOVA technique is performed for the energy of the CH_4 molecule in Table 4 to illustrate a summary of the information obtained. Again, recall that the factorial design in Table 2 describes the values for all the MEAM parameters for the 64 different simulations (treatments) of the CH_4 molecule. The CH_4 molecule is

minimized using a gradient-based energy minimization technique, and the final energy and positions of the atoms are used for various responses: atomization energy, C–H bond length, and H–C–H bond angle. The data from the ANOVA procedure are listed in Table 4, which includes several columns of data: 1) the total sum of squares as well as the sum of squares attributed to each factor and the error; 2) the degrees of freedom (d.f.), which is equal to 1 for all factors; 3) the mean squares of the various factors and the error, which is an unbiased estimate of the variance; 4) the F test statistic, which is a ratio of the mean square of the factors to that of the error; and 5) the p-value, which is related to the hypothesis that there are differences in the variance due to a particular factor. There are a few important results that can be gleaned from this information. First, the p-value indicates which factors are significant; a value below 0.05 indicates a 95% confidence level for significance. Hence, in terms of the energy of this molecule, there are 7 statistically significant factors at a 95% confidence level: esub1, asub1, b02, alat2, esub2, deltas12, and res12. Second, the sum of squares column indicates the degree of influence of each parameter and dividing by the total sum of squares gives the percent contribution for each factor to the response. So, the percent contribution of each statistically significant factor can be calculated: 49.7% (deltas12), 22.1% (esub2), 17.5% (esub1), and 9.1% (asub1) (i.e., 98.3% of the variability in this study is caused by these 4 factors). Another way of interpreting this is that a linear regression model is fit using the various parameters outlined in Table 4, and the sum of squares column is simply a reflection of how much of the original variability in properties is accounted for by each parameter. In this manner, the error term gives the ability to describe the data using this linear regression model; in truth, if a predictive model is the goal, then including parameter interaction terms or other mathematical model formulations may better capture the relationship between parameters and responses. This can be easily accomplished but is outside the scope of this work.

Table 4 ANOVA analysis of energy of CH_4 molecule

Source	Sum Sq.	d.f.	Mean Sq.	F	p-value (Prob$>F$)
meam.alpha1	0.000	1	0.000	0.000	1.000
meam.b01	0.000	1	0.000	0.000	0.999
meam.b11	0.000	1	0.000	0.000	0.998
meam.b21	0.000	1	0.000	0.000	0.992
meam.b31	0.000	1	0.000	0.000	0.988
meam.alat1	0.000	1	0.000	0.000	0.987
meam.esub1	31.065	1	31.065	16,413.814	0.000
meam.asub1	16.099	1	16.099	8,506.316	0.000
meam.alpha2	0.000	1	0.000	0.005	0.946
meam.b02	0.016	1	0.016	8.591	0.006
meam.b12	0.001	1	0.001	0.717	0.402
meam.b22	0.002	1	0.002	0.813	0.373
meam.b32	0.000	1	0.000	0.036	0.850
meam.alat2	1.895	1	1.895	1,001.323	0.000
meam.esub2	39.280	1	39.280	20,754.279	0.000
meam.asub2	0.000	1	0.000	0.000	0.991
meam.rozero2	0.000	1	0.000	0.004	0.950
parameter.deltas12	88.377	1	88.377	46,695.244	0.000
parameter.res12	0.997	1	0.997	526.704	0.000
parameter.alphas12	0.000	1	0.000	0.086	0.771
parameter.attrac11	0.000	1	0.000	0.004	0.949
parameter.repuls11	0.000	1	0.000	0.000	0.999
parameter.attrac12	0.000	1	0.000	0.113	0.739
parameter.repuls12	0.000	1	0.000	0.000	0.998
parameter.attrac22	0.000	1	0.000	0.000	0.999
parameter.repuls22	0.000	1	0.000	0.000	0.999
Error	0.070	37	0.002	...	...
Total	177.803	63	...	...	...

3.2 Interatomic Potential Parameter Sensitivity via Analysis of Variance

The present study contains 30 different tests, which each have a number of different responses that can be evaluated against the factors explored in the factorial design using an ANOVA procedure similar to that illustrated in Table 4. The large number of responses included in the present study required exploring different techniques for visualizing and analyzing the information. One way to examine this informa-

tion is to use a contour plot to visualize how the percent contribution for each factor changes for each molecule. This type of plot is shown in Fig. 3, which depicts how the percent contribution to the energy for each factor changes as a function of molecule type. On the horizontal axis are the 30 different molecules in the present study, ranging from C_2 to C_2H_5. On the vertical axis are the 26 different MEAM parameters. The contours are related to the percent contribution. The bands of increased percent contribution indicate the parameter sensitivity for the properties (in this case, molecule energy) (i.e., visually depicting the relative influence of parameters for tuning a particular property). In general, the deltas12 term (E^0_{CH}) has the largest influence for most molecules, followed by esub1 (E^0_C) and esub2 (E^0_H), as might be expected from the formulation of the potential (i.e., these parameters define the sublimation/cohesive energies of the reference state for C, H, and C–H). The bands indicate that these parameters affect the molecule energies in a similar manner despite the very different geometries associated with the different molecules. The C and H dimers are heavily influenced by asub1 (A_C) and esub2 (E^0_H). The fact that the H dimer energy is completely influenced by E^0_H is no surprise; the reference state for the H MEAM potential is the dimer. However, interestingly, the C dimer energy is influenced much more by A_C (scaling factor for the background electron density of C, Eq. 2) rather than E^0_C, which is due to the diamond cubic reference state for C as opposed to the dimer reference state.

The data displayed on the subsequent contour plots were computed in the same manner, but with bond distances (Figs. 4 and 5), bond angles (Figs. 6–8), dihedral angles (Figs. 9–11), and the energy vs. distance curves for various molecules and paths (Figs. 12 and 13).

Figures 4 and 5 show the contribution of the MEAM parameters to the C–H and C–C bond distances, respectively. The bond distances are most influenced by res12 (R^0_{CH}), asub1 (A_C), and alat1 (R^0_C). The C–H bond distance is affected by R^0_{CH}, a parameter associated with the distance of the C–H reference state (dimer), and A_C, the scaling factor for the electron density. The C–C bond distance is mainly affected by R^0_C for all molecules with a minor effect of A_C for C_2 and C_2H_2.

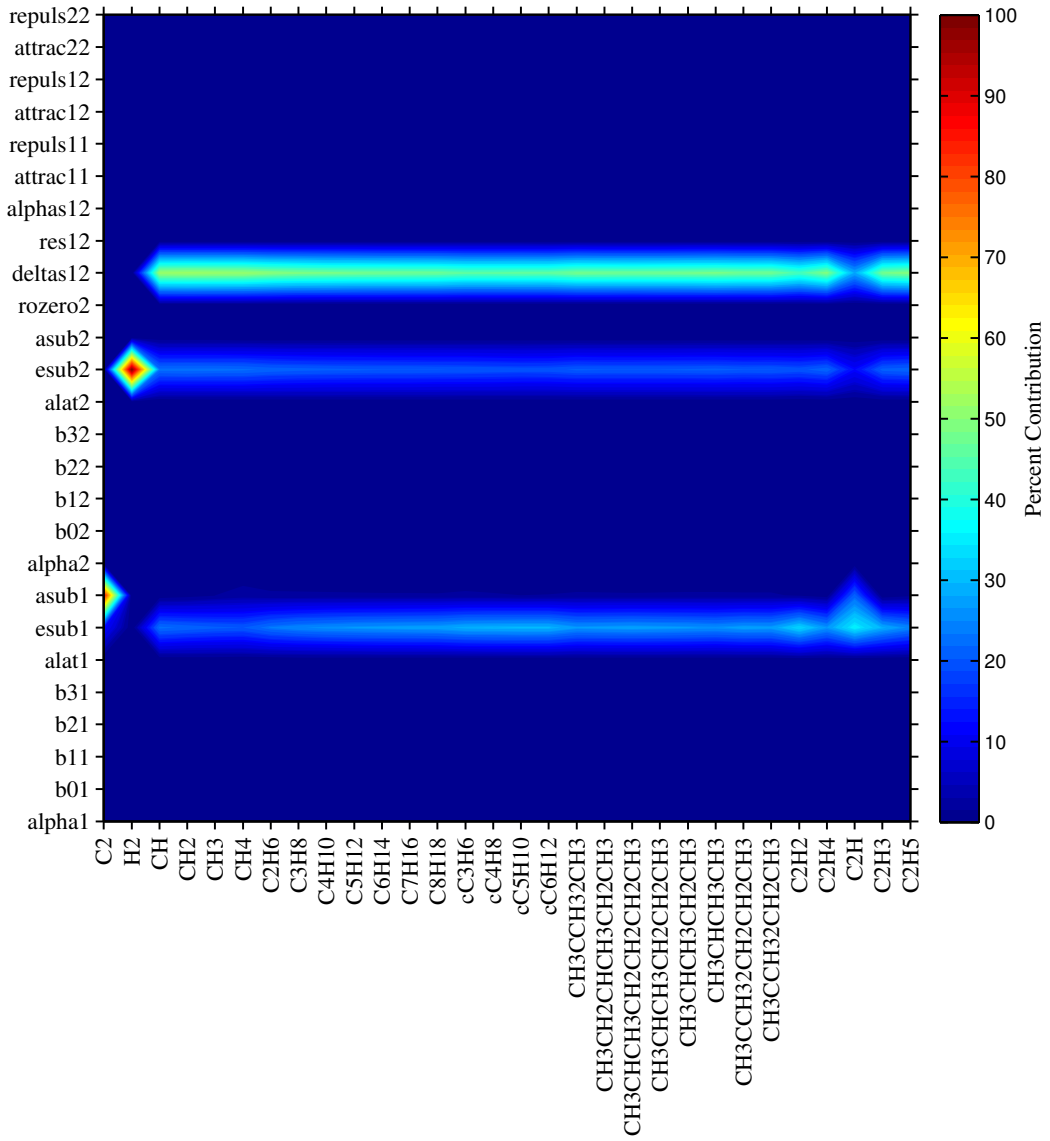

Fig. 3 Summary of ANOVA results for the energies of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule.

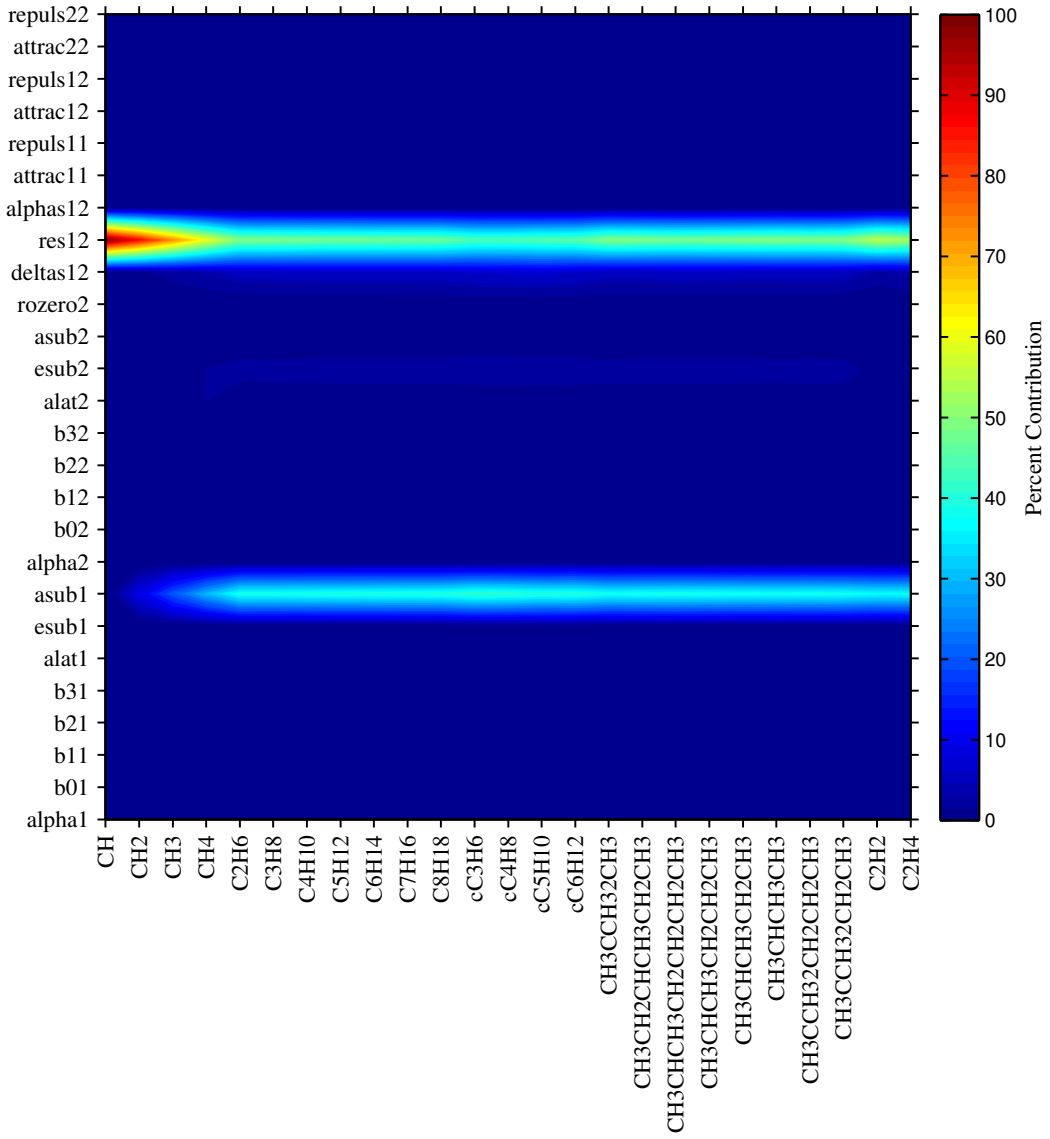

Fig. 4 Summary of ANOVA results for the C–H bond distances of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond distance for all C–H bonds was used for those molecules with multiple C–H bonds.

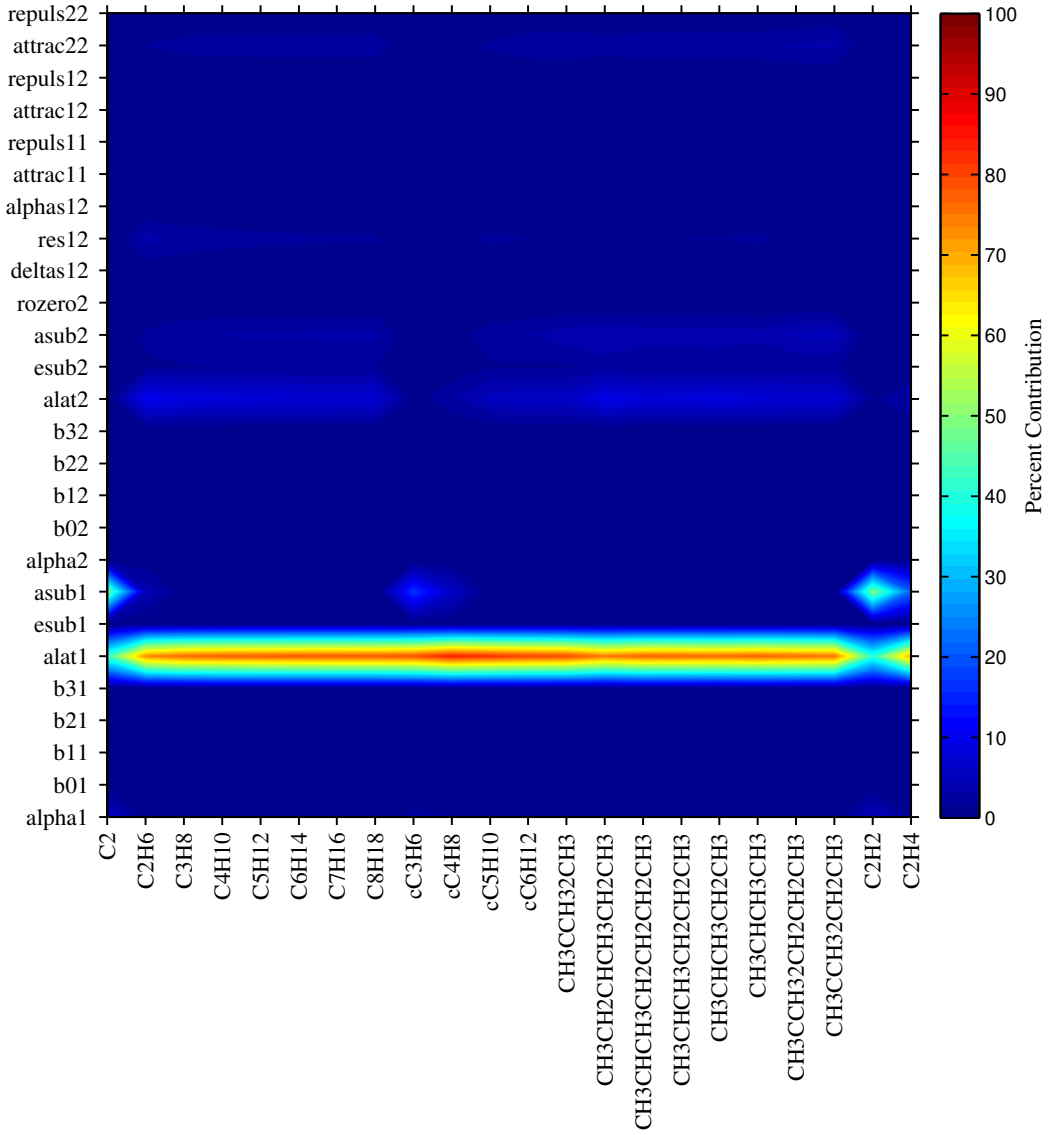

Fig. 5 Summary of ANOVA results for the C–C bond distances of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond distance for all C–C bonds was used for those molecules with multiple C–C bonds.

Figures 6–8 show the contribution of the MEAM parameters to the C–C–C, C–C–H, and H–C–H bond angles, respectively. In contrast to the energies and bond distances of the various molecules, multiple MEAM parameters contribute to the different bond angles and these contributions change as a function of the molecule type. In general, the bond angles are mainly influenced by asub1 (A_C), alat1 (R_C^0), attrac22 (δ_H^a), asub2 (A_H), alat2 (R_H^0), b12 ($\beta_H^{(1)}$), and res12 (R_{CH}^0). Interestingly, the relationship between MEAM parameters and the bond angles are not always as intuitive. For instance, the variation in the C–C–C bond angle was mainly influenced by modifications to the H MEAM parameters (i.e., R_H^0 and δ_H^a), which also played a significant role in the C–C–H and H–C–H bond angles. There are also significant shifts between different molecule families (alkanes vs. cycloalkanes vs. polyolefins) in terms of the significance of these various MEAM parameters. That is, the effects to these bond angles in one molecule is not necessarily correlated to effects in another molecule; this is further examined by correlating properties for all the various molecules.

Figures 9–11 show the contribution of the MEAM parameters to the C–C–C–C, C–C–C–H, and H–C–C–H dihedral angles, respectively. Similarly to the bond angles, the dihedral angles are affected by a number of the same MEAM parameters: asub1 (A_C), alat1 (R_C^0), attrac22 (δ_H^a), asub2 (A_H), alat2 (R_H^0), b12 ($\beta_H^{(1)}$), and res12 (R_{CH}^0). Again, interestingly, some of the strongest influences to the C–C–C–C dihedral angles are the contributions of the single element H MEAM potential (δ_H^a, A_H, $\beta_H^{(1)}$) in addition to some of the single element C and C–H interaction MEAM parameters. In general, the attraction term for hydrogen δ_H^a has a large influence on all of the dihedral angles. Some of the trends in MEAM parameter sensitivities change with molecule as well. For instance, the H–C–C–H dihedral angle is significantly affected by R_C^0 for the alkane series of molecules with greater than 3 C atoms.

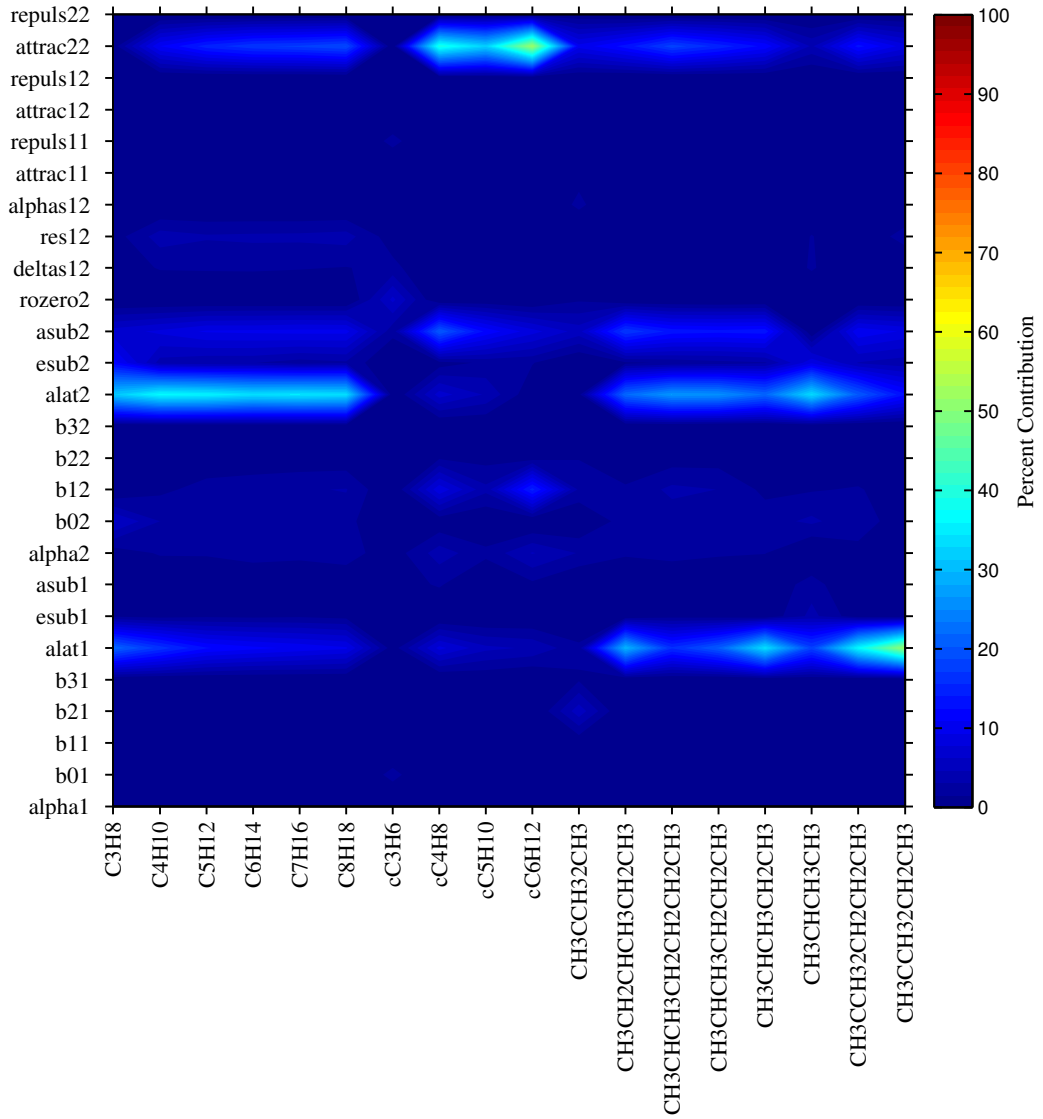

Fig. 6 Summary of ANOVA results for the C–C–C bond angles of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond angle for all C–C–C bond angles was used for those molecules with multiple C–C–C bond angles.

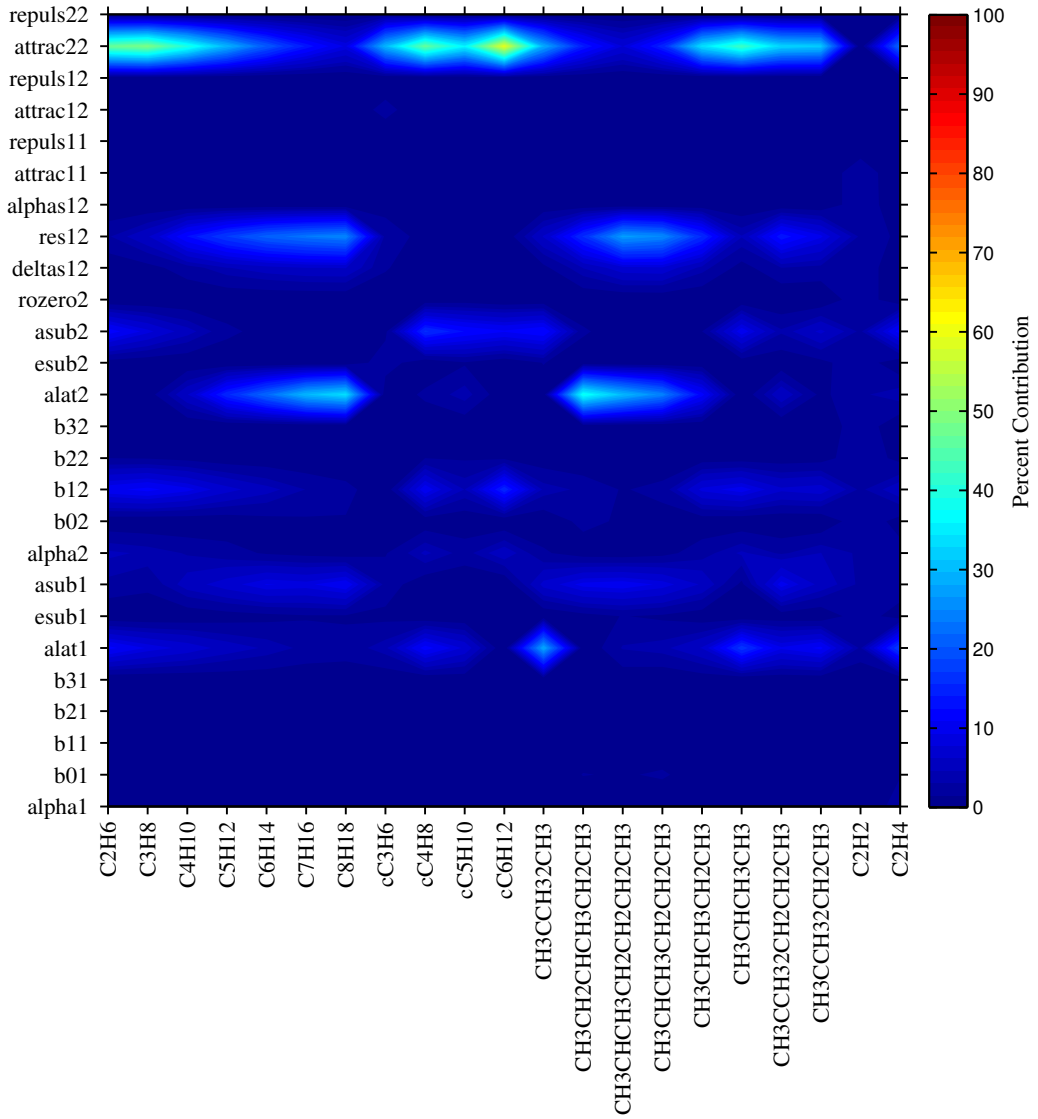

Fig. 7 Summary of ANOVA results for the C–C–H bond angles of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond angle for all C–C–H bond angles was used for those molecules with multiple C–C–H bond angles.

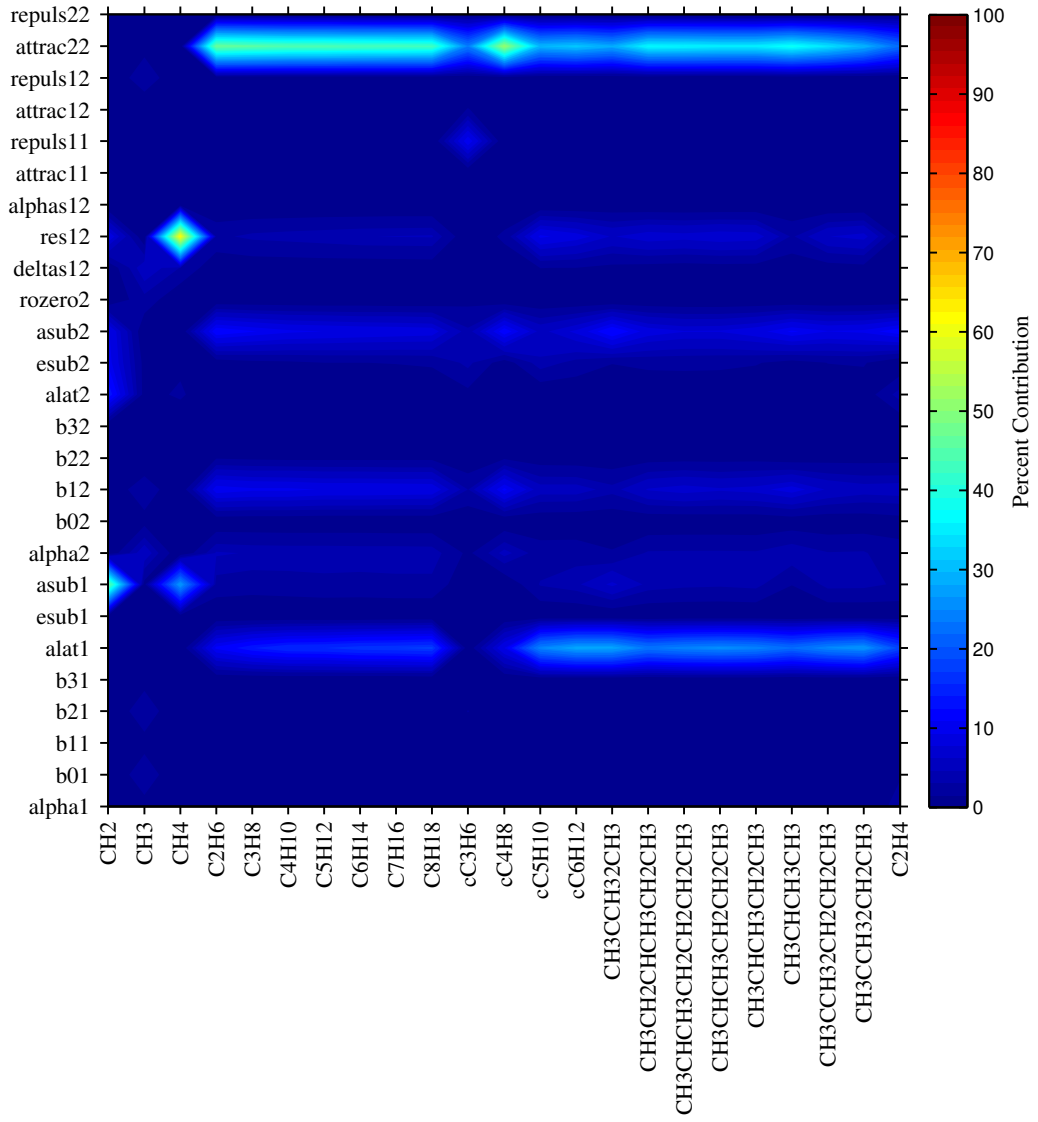

Fig. 8 Summary of ANOVA results for the H–C–H bond angles of all molecules examined. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean bond angle for all H–C–H bond angles was used for those molecules with multiple H–C–H bond angles.

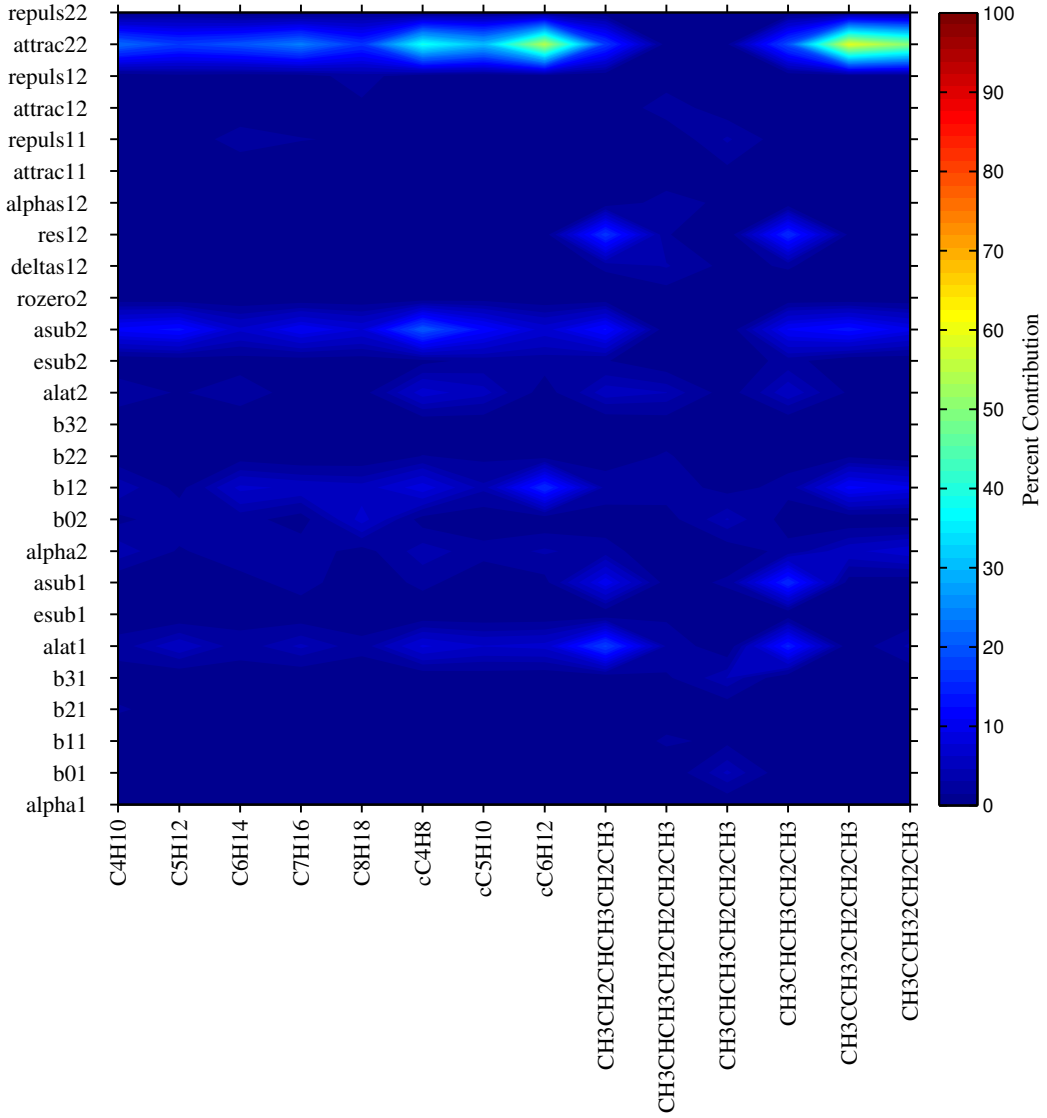

Fig. 9 Summary of ANOVA results for the C–C–C–C dihedral angles of all molecules examined. Since there are primarily 2 dihedral angles in the tested molecules (60° and 180°), the absolute value of the minimum difference between each dihedral angle and these 2 reference angles was used. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean dihedral angle for all C–C–C–C angles was used for those molecules with multiple C–C–C–C dihedral angles.

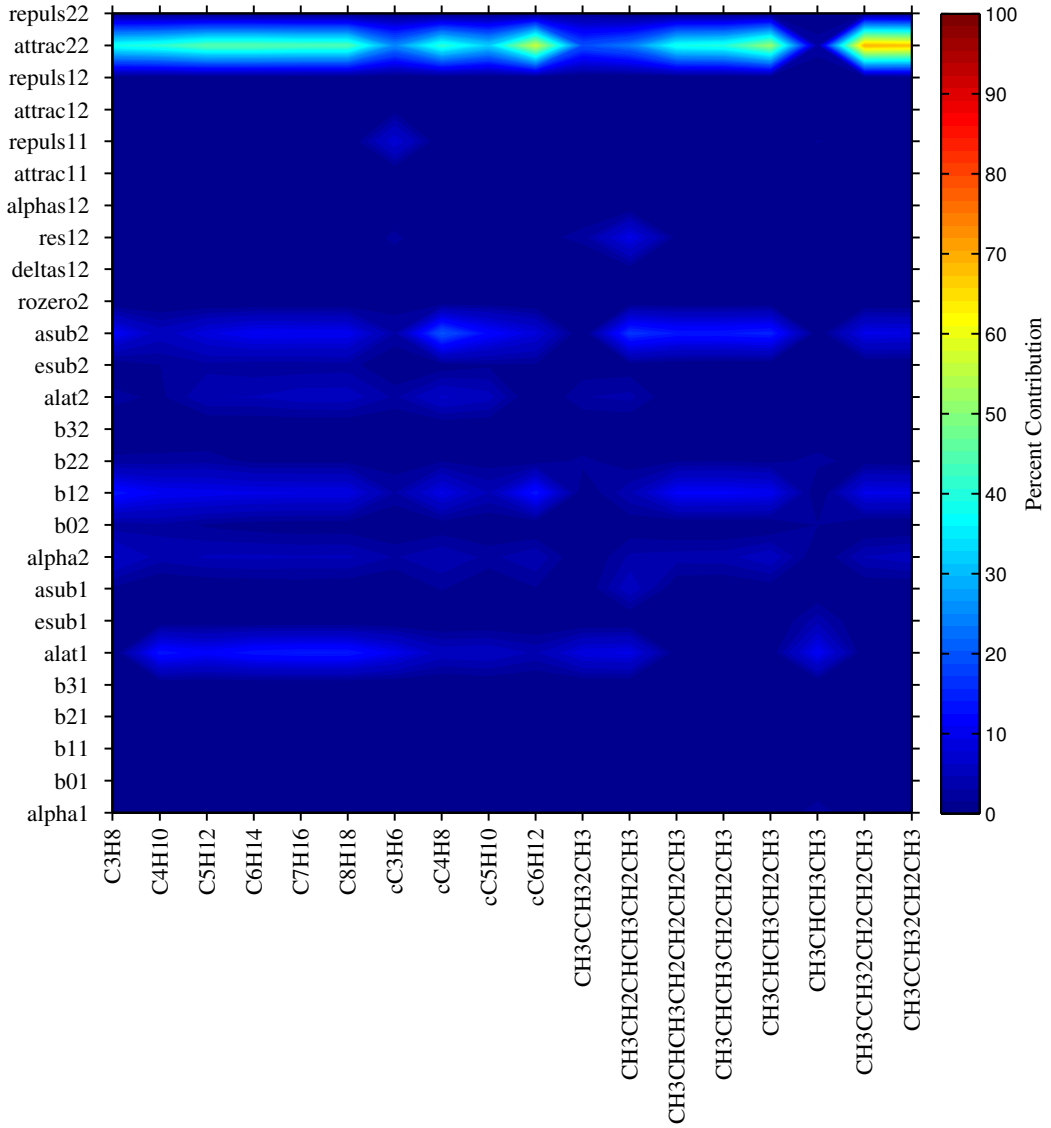

Fig. 10 Summary of ANOVA results for the C–C–C–H dihedral angles of all molecules examined. Since there are primarily 2 dihedral angles in the tested molecules (60° and 180°), the absolute value of the minimum difference between each dihedral angle and these 2 reference angles was used. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean dihedral angle for all C–C–C–H angles was used for those molecules with multiple C–C–C–H dihedral angles.

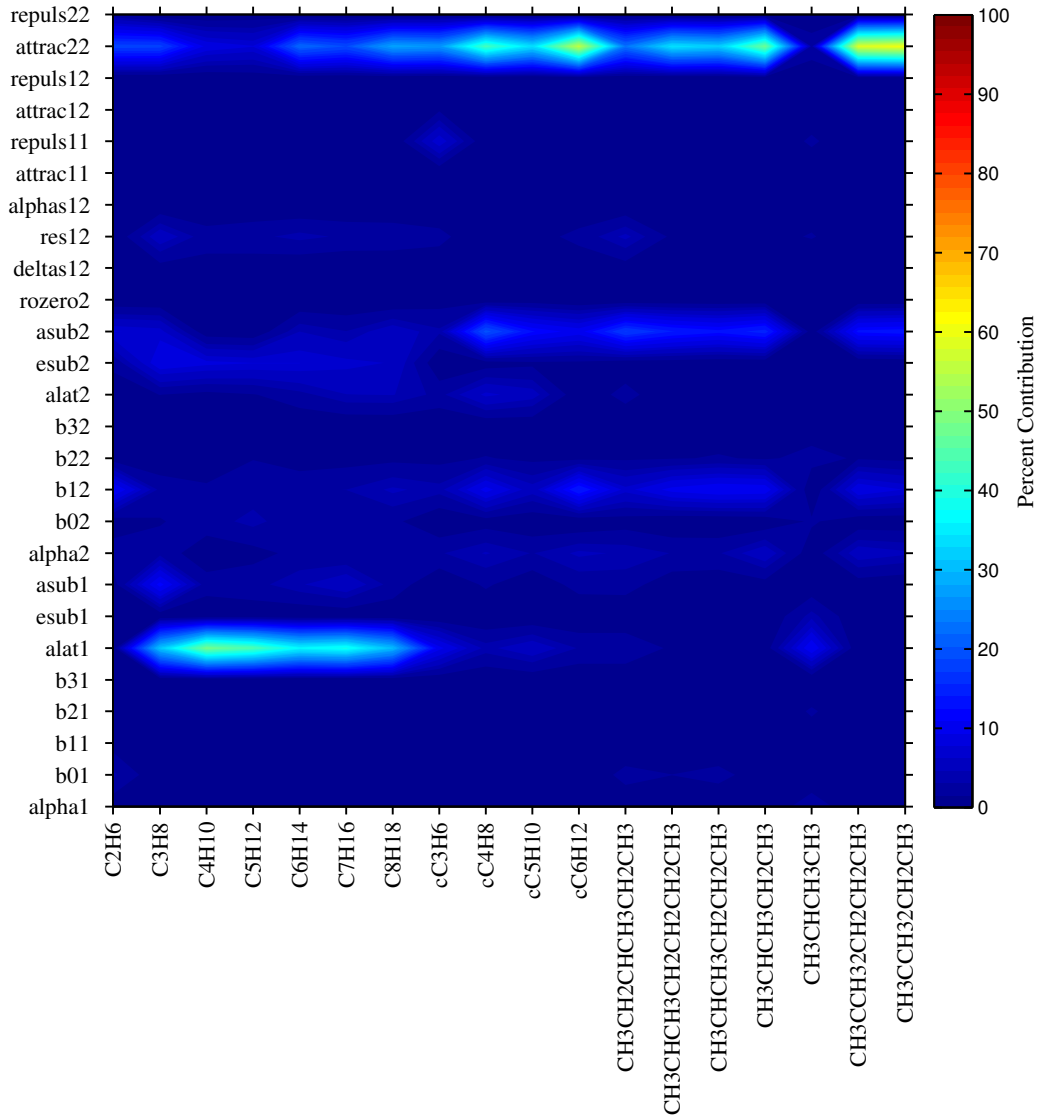

Fig. 11 Summary of ANOVA results for the H–C–C–H dihedral angles of all molecules examined. Since there are primarily 2 dihedral angles in the tested molecules (60° and 180°), the absolute value of the minimum difference between each dihedral angle and these 2 reference angles was used. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule. The mean dihedral angle for all H–C–C–H angles was used for those molecules with multiple H–C–C–H dihedral angles.

Figures 12 and 13 show the contribution of the MEAM parameters to the energy vs. distance relationships for various molecules and paths, as outlined in Nouranian et al.[39] These figures compare the MEAM-computed energy vs. distance plots to the same first principle energy vs. distance plots for separation between various molecules (C_2H_6, CH_4, and H_2) and atoms (C–C, H–H, C–H). As a measure of how similar the 2 curves are, the mean absolute error and the root mean square error (i.e., L_1 and L_2 norms, respectively) are used as the similarity/distance metrics within the analysis. In general, the same parameters are important for both metrics. Interestingly, the MEAM parameters that are mainly driving these energy-distance metrics are related to the single element H MEAM potential: attrac22 (δ_H^a), asub2 (A_H), esub2 (E_H^0), alat2 (R_H^0), and b12 ($\beta_H^{(1)}$). For the hydrogen dimers energy-distance relationships, this is to be expected, but this also shows that the interaction between different molecules (i.e., van der Waals' interaction) is largely influenced by the H atoms on the C_2H_6 and CH_4 molecules. Last, the percent contribution of the MEAM parameters to the responses does not always sum to 100% (i.e., Fig. 13); in some cases, the residuals (or error term in Table 4) for the linear regression model applied during the ANOVA analysis are a significant fraction of the overall variability of the responses.

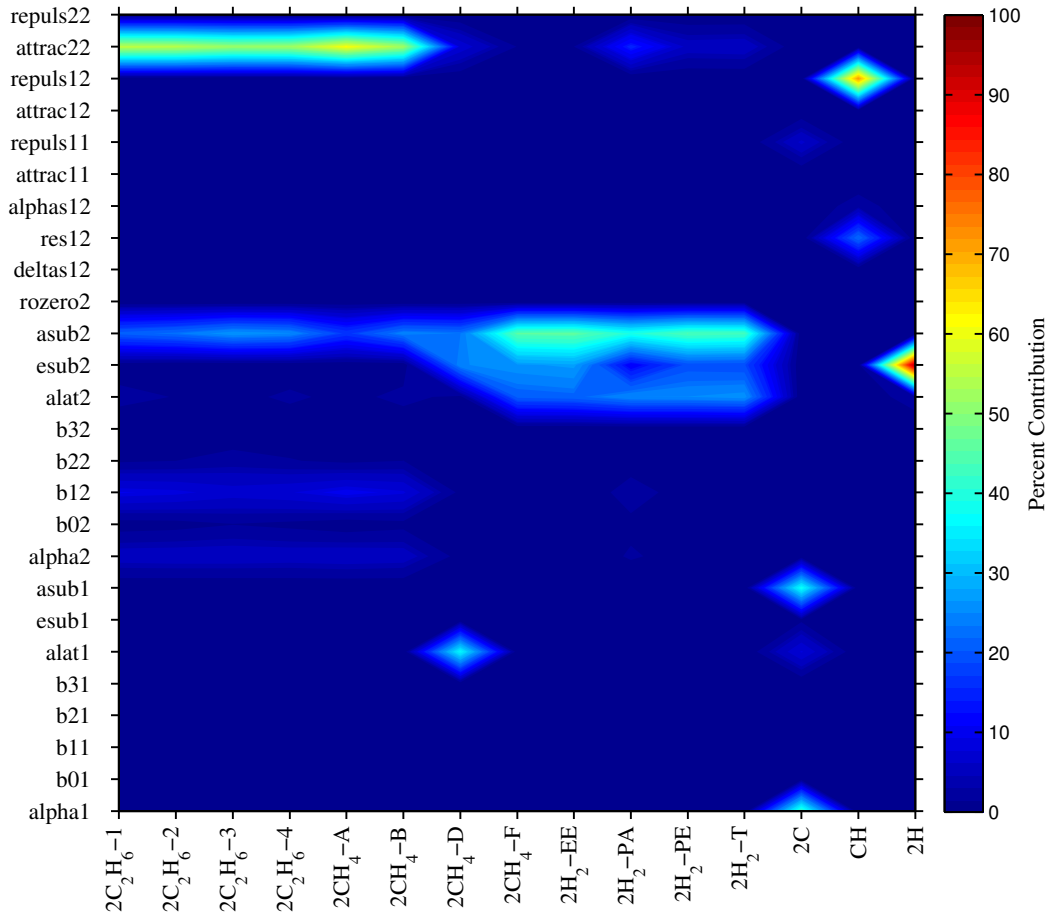

Fig. 12 Summary of ANOVA results for the energy vs. distance relationships for various molecules and paths. In this contour plot, the mean absolute error between the calculated MEAM curve and first principles data at the same distances was used as a response variable. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule.

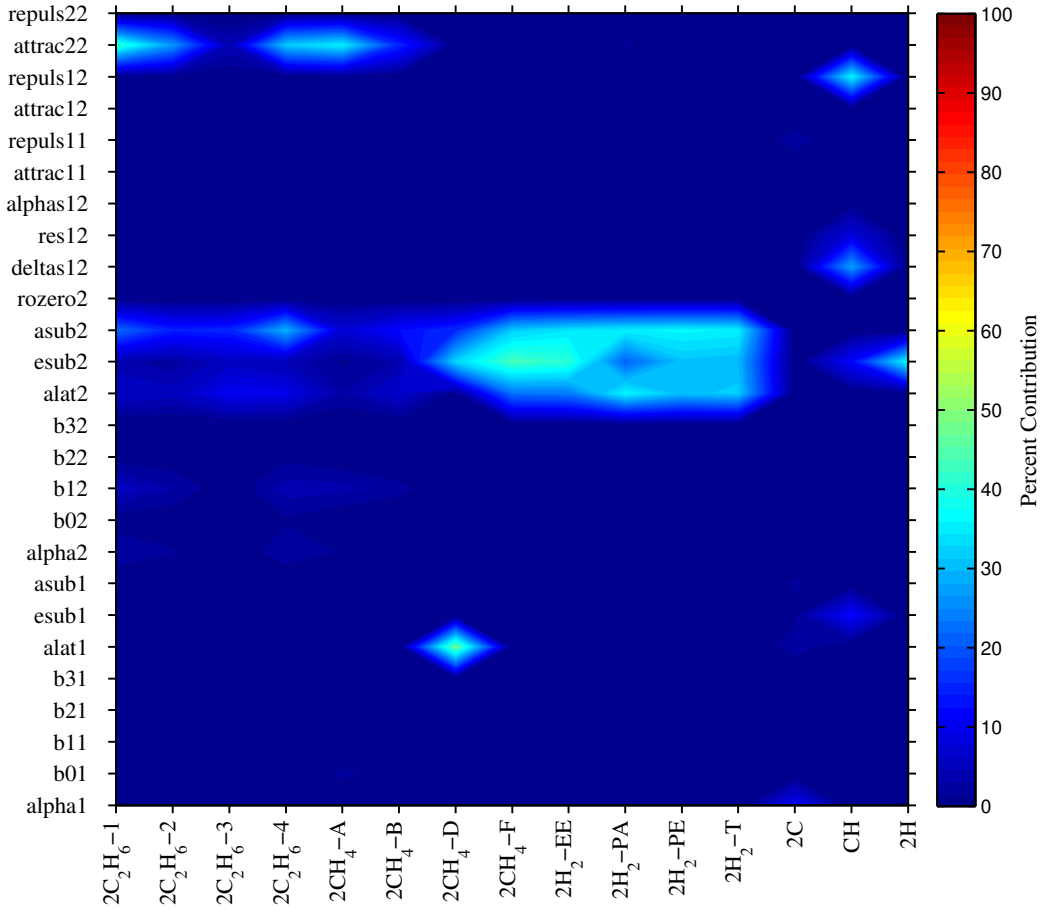

Fig. 13 Summary of ANOVA results for the energy vs. distance relationships for various molecules and paths. In this contour plot, the root mean square error between the calculated MEAM curve and first principles data at the same distances was used as a response variable. The percent contribution by each of the MEAM parameters was determined by dividing the sum of squares attributed to each parameter by the total sum of squares (e.g., see Table 4), which was repeated for each molecule.

In general, these analyses can be used in a few ways in the potential development process. First, Figs. 3–13 can help to evaluate the parameter sensitivity for various properties in the training set. Ultimately there will be tradeoffs that are encountered and can be managed through a multi-objective optimization approach, but this can help identify which parameters may be involved with optimization of particular properties. In some cases, depending on the level of complexity of the interatomic potential, it may be easier to spawn multiple suboptimization problem formulations for various properties rather than trying to optimize all properties for all parameters. This can also be used for defining the appropriate parameter ranges for potential development. While Figs. 3–13 show how important different parameters are for different properties and molecules, it also shows which properties do not significantly influence properties. This analysis may provide guidance in terms of expanding the range of various parameters to assess whether the property in question is truly unaffected by the parameter or whether the parameter range needs to be expanded.

3.3 Property Similarity for Different Molecules/Potentials

The present study contains properties for 30 different molecules as a function of different combinations of MEAM parameters. The similarity between the properties of different molecules can shed some light on the uniqueness of different molecules for potential use when constructing training sets for interatomic potentials. In this particular study, a Latin hypercube sampling technique was used to sample the 26-dimensional parameter space described for the fractional factorial design. From this, 1000 different MEAM potentials were each used to generate the 30 different molecules (i.e., 30,000 molecules) and their corresponding properties. The Pearson correlation coefficient R was used as a metric for similarity of the different properties (stemming from the 1000 different MEAM potentials). Figures 14–20 show the Pearson correlation coefficient for different properties (energies, bond distances, and bond angles) when compared to all other molecules. For instance, the Pearson correlation coefficients of the energies of the molecules for the Latin hypercube design is shown in Fig. 14. A high correlation coefficient indicates that the energy of the 2 molecules on the horizontal and vertical axes scales similarly even as the MEAM parameters are changed for the 1000 different MEAM potentials. For example, if a new MEAM interatomic potential is chosen and the energy of C_3H_8 ($E_{C_3H_8}$) increases (by δE_1), the energy of C_8H_{18} ($E_{C_8H_{18}}$) will also increase (by δE_2) in such a manner that the slope of the increases is con-

stant and predictable (i.e., $E_{C_8H_{18}} = (\delta E_2/\delta E_1)\, E_{C_3H_8}$). While this is perhaps not unexpected for the different molecules within the alkane series, for instance, it is interesting that there is large correlation ($R = 1$) between different families of molecules (e.g., between C_8H_{18} and cycloalkane cC_6H_{12}). A low (or no) correlation indicates that the properties for 2 molecules are not in line with each other; the energy of C_2 does not align with the energy of H_2 (not surprising since different parts of the interatomic potential control these properties independent of one another). While this low (or no) correlation for 2 molecules is in itself not of high value, the rows/columns of low/no correlation can shed light on the uniqueness of certain molecules. For instance, if H_2 dimer is not in a training set for the potential, then one would not expect any of the alkane, cycloalkane, or polyolefin molecules to accurately capture this response. As has been shown in Fig. 3, this particular situation arises from different potential parameters controlling the response of these molecules. Figure 15 also shows the correlation coefficient for the same properties for the fractional factorial design (i.e., 64 MEAM potentials vs. the 1000 MEAM potentials in the Latin hypercube design). In general, the fractional factorial design yields similar behavior to the more expansive coverage offered by the 1000-point Latin hypercube sampling strategy, with a fraction of the simulations.

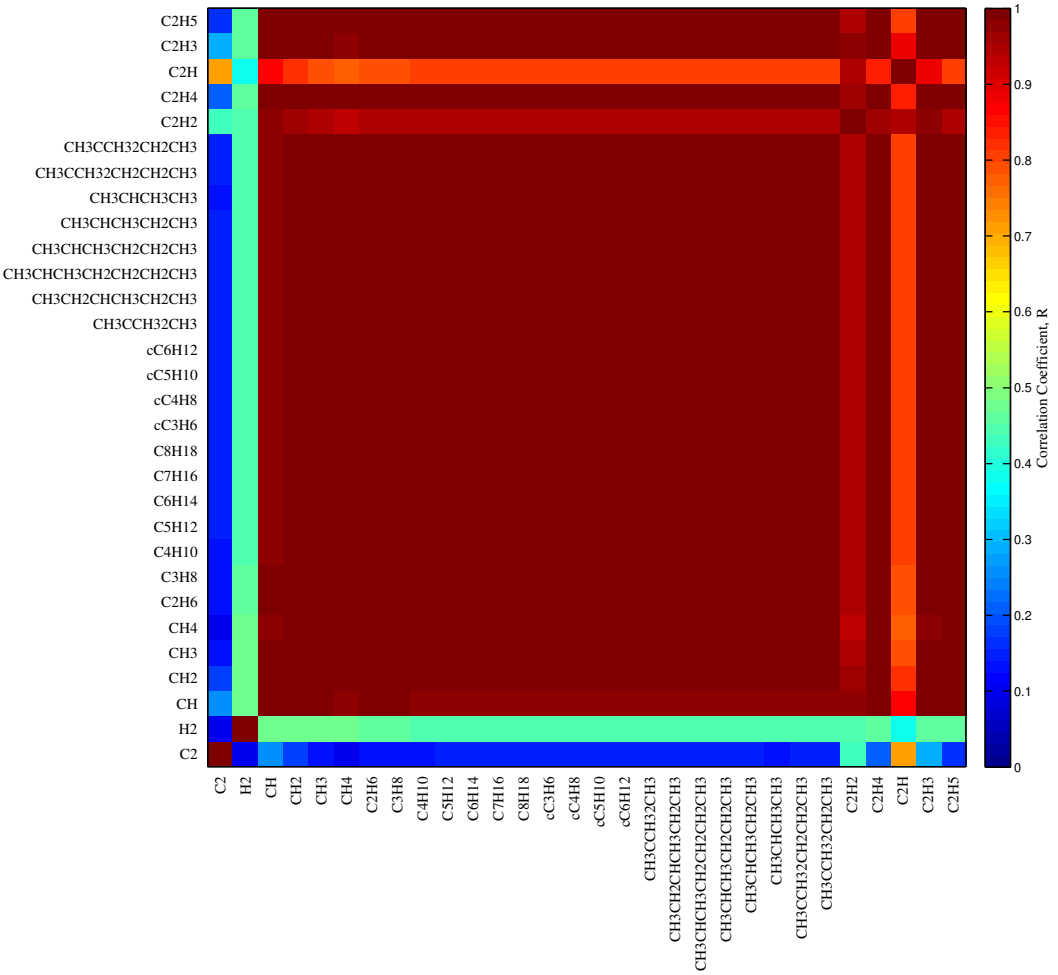

Fig. 14 Correlation coefficient R for the energies of all molecules examined (for the Latin hypercube sampling design). In this plot, each response is plotted against all of the other responses. A high correlation coefficient indicates the strength of the linear dependence between the 2 responses (i.e., a high correlation, R=1, for responses plotted against themselves).

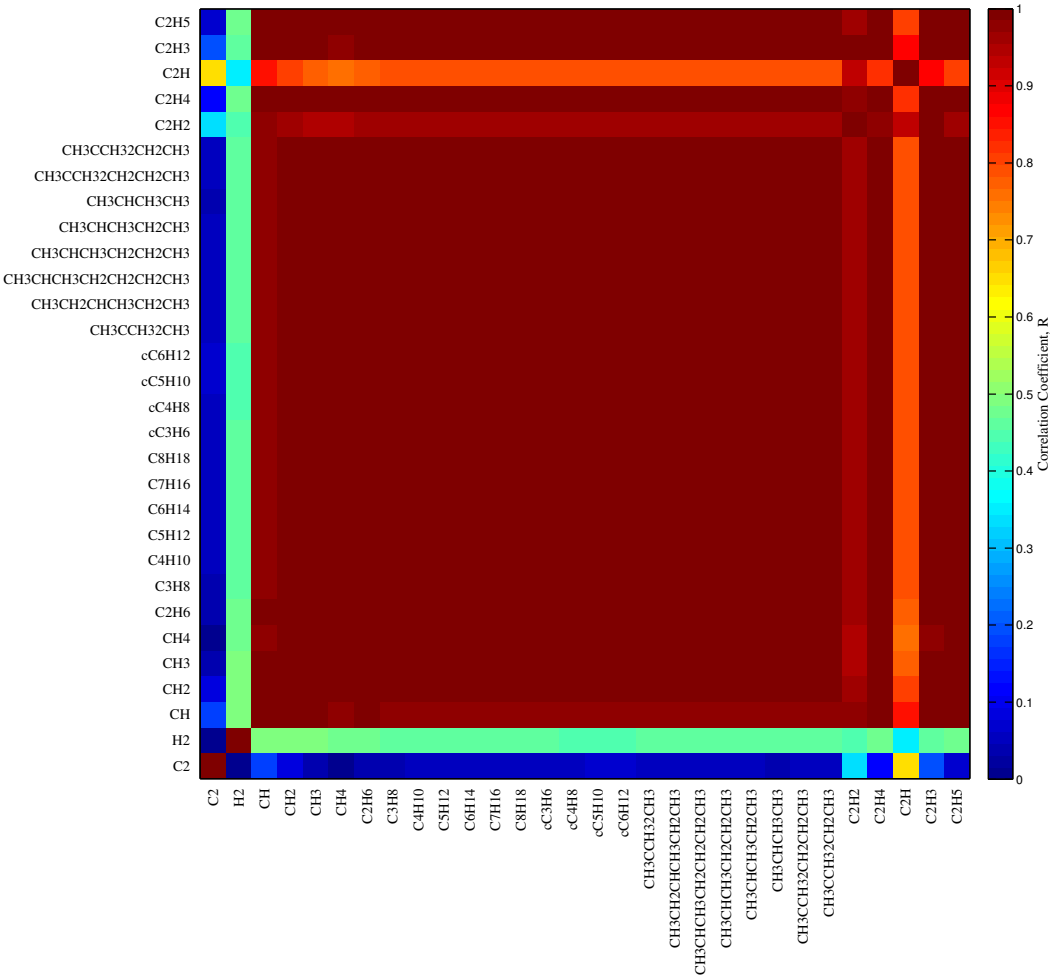

Fig. 15 Correlation coefficient R for the energies of all molecules examined (for the fractional factorial design case). In this plot, each response is plotted against all of the other responses. A high correlation coefficient indicates the strength of the linear dependence between the 2 responses (i.e., a high correlation, R=1, for responses plotted against themselves).

This correlation between properties is also plotted for bond distances (Figs. 16 and 17) and bond angles (Figs. 18–20). In terms of the C–H bond distance (Fig. 16), this is the most correlated properties between the different molecules (i.e., this property behaves similarly with respect to the interatomic potential no matter what the molecule is, the one exception being the C–H dimer bond distance). On the other hand, the C–C bond distance (Fig. 17) is very different in terms of its correlation between molecules. This plot indicates that C_2, cC_3H_6, cC_4H_8, C_2H_2, and C_2H_4 are very dissimilar from the remaining molecules, which tend to have C–C bond distances that are similarly affected by interatomic potential. This is perhaps expected; the dimers (C_2 in this case) tend to have very different properties than the other molecules, the cycloalkanes with C less than 4 (cC_3H_6, cC_4H_8) have constrained bond angles that can alter their properties, and the alkene/alkyne molecules (C_2H_2, C_2H_4) have a very different bonding state between the C atoms due to the absence of H atoms. Interestingly, comparing Fig. 17 with parameter plot in Fig. 5, one notices subtle changes in the percent contribution of the different MEAM parameters to the C–C bond distance of these molecules; this translates into very noticeable deviations from perfect correlation in Fig. 17. Again, this would be as expected. For instance, while asub1 (A_C) affects the C–C distance of these molecules, this parameter has little-to-no effect on the other molecules. If held fixed, the correlation will be higher; if allowed to change, this will affect the properties of only a select group of molecules, resulting in a lower correlation as observed herein.

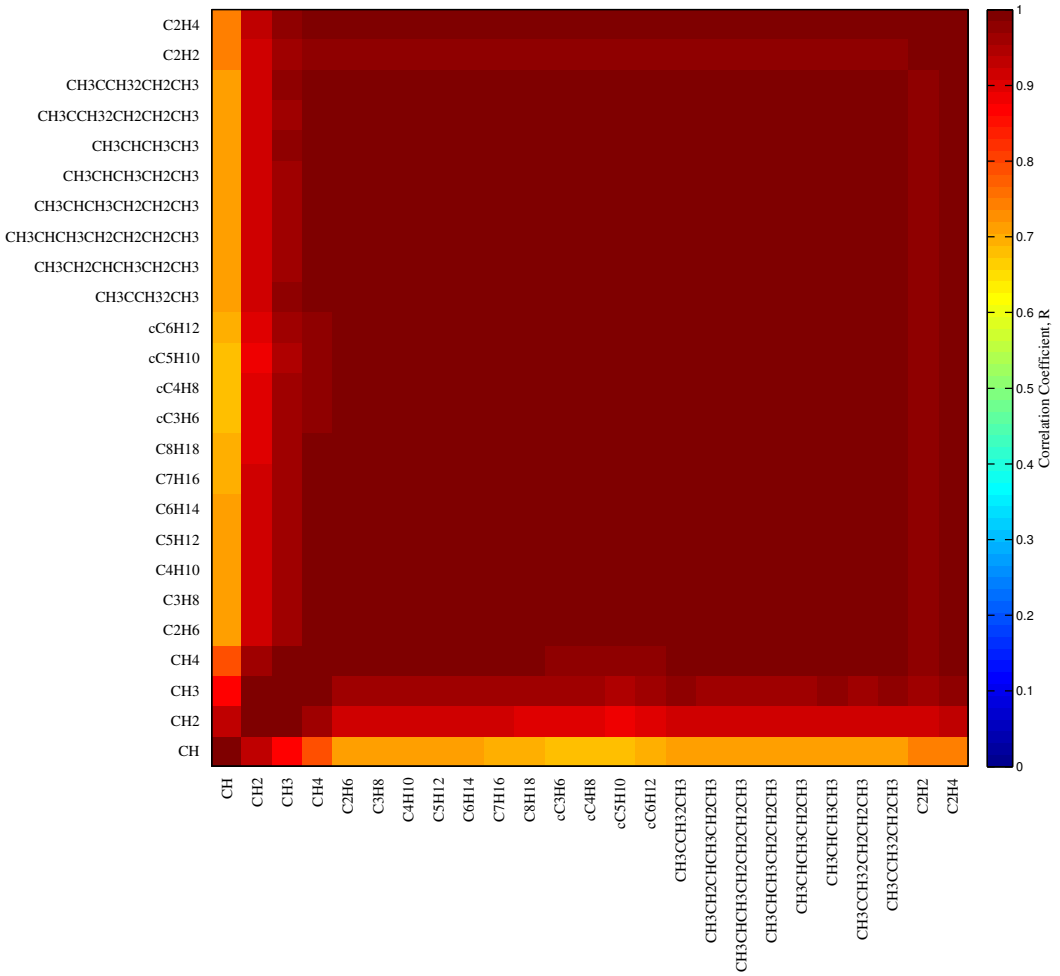

Fig. 16 Correlation coefficient R for the C–H bond distances of all molecules examined. The mean bond distance for all C–H bonds was used for those molecules with multiple C–H bonds.

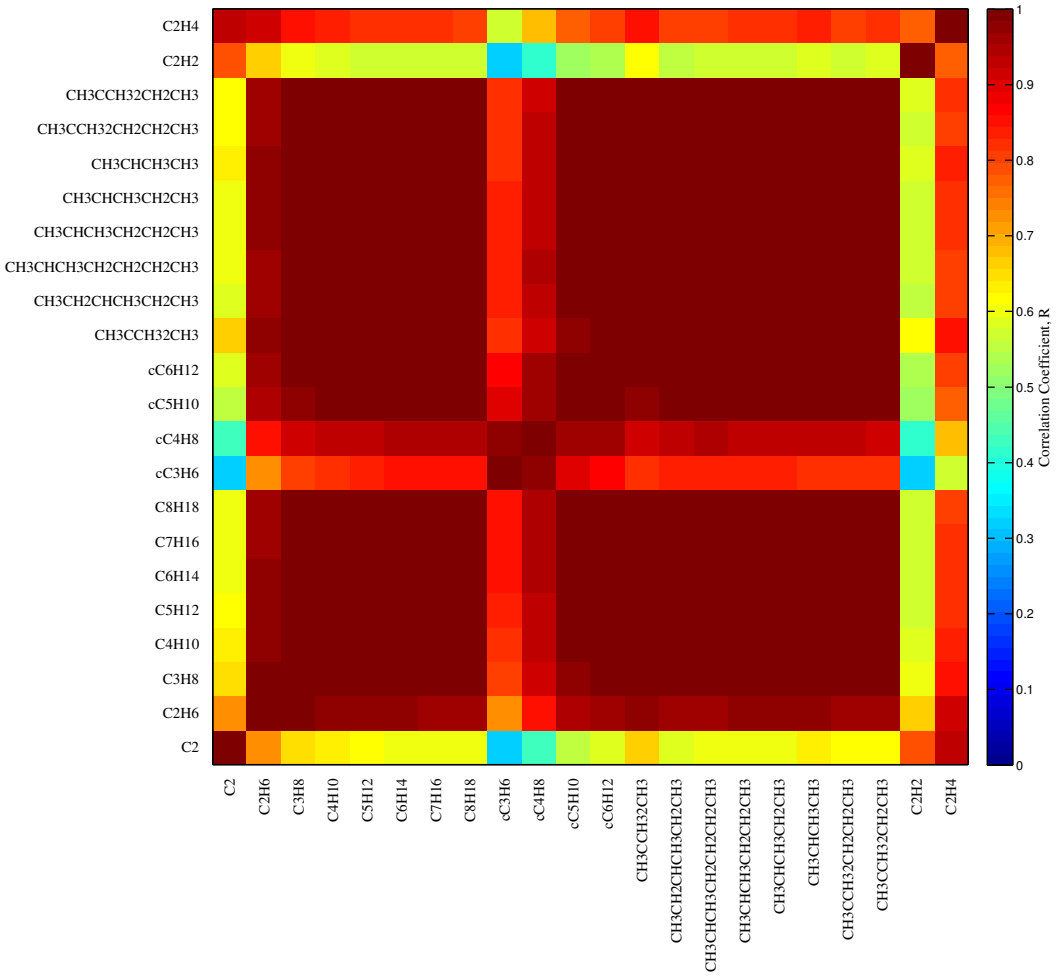

Fig. 17 Correlation coefficient R for the C–C bond distances of all molecules examined. The mean bond distance for all C–C bonds was used for those molecules with multiple C–C bonds.

Figures 18–20 show the correlation between molecules with respect to the C–C–C, C–C–H, and H–C–H bond angles, respectively. In these cases, there are definitive dissimilarities between different molecules. For instance, the cycloalkanes are dissimilar from other molecules in terms of the C–C–C bond angles, but have varying degrees of similarity in terms of the C–C–H bond angles and H–C–H bond angles. Similar observations can be made about some of the other molecules and families of molecules. In part, the dissimilarity between molecules, particularly present in Fig. 19 for C–C–H bond angles, can also be driven by the calculation of the bond angle itself. The bond angles represented in this work are the result of taking all bond angles of a particular type and averaging the values, thus different local environments (e.g., head or tail group of a linear alkane molecule) can influence the average bond angle. For example, even a simple molecule such as the C_6H_{14} alkane molecule has 19 bonds, 36 bond angles, and 45 dihedral angles, which are broken down in the following manner: 5 C–C bonds, 14 C–H bonds, 4 C–C–C bond angles, 22 C–C–H bond angles, 10 H–C–H bond angles, 3 C–C–C–C dihedral angles, 18 C–C–C–H dihedral angles, and 24 H–C–C–H dihedral angles. Of the 22 C–C–H bond angles, most bond angles are within 0.2° of each other (≈109.7°) except for 2 bond angles at the head and tail of the molecule that have bond angles of 113.28°. However, it was necessary to reduce the dimension of this data in some manner and the average response was deemed an appropriate measure for the present study (e.g., 1000 potentials, 1 C_8H_{18} molecule, 30 different C–C–H bond angles).

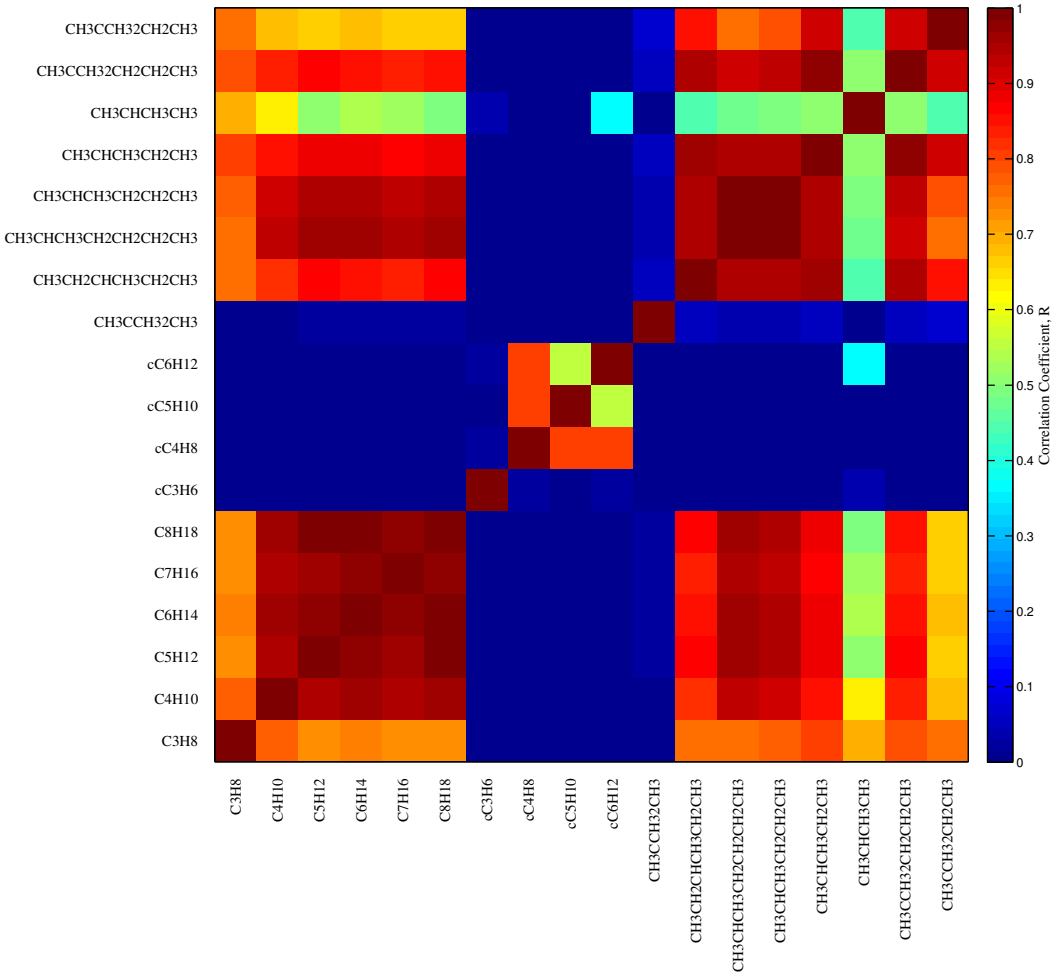

Fig. 18 Correlation coefficient R for the C–C–C bond angles of all molecules examined. The mean bond angle for all C–C–C bond angles was used for those molecules with multiple C–C–C bond angles.

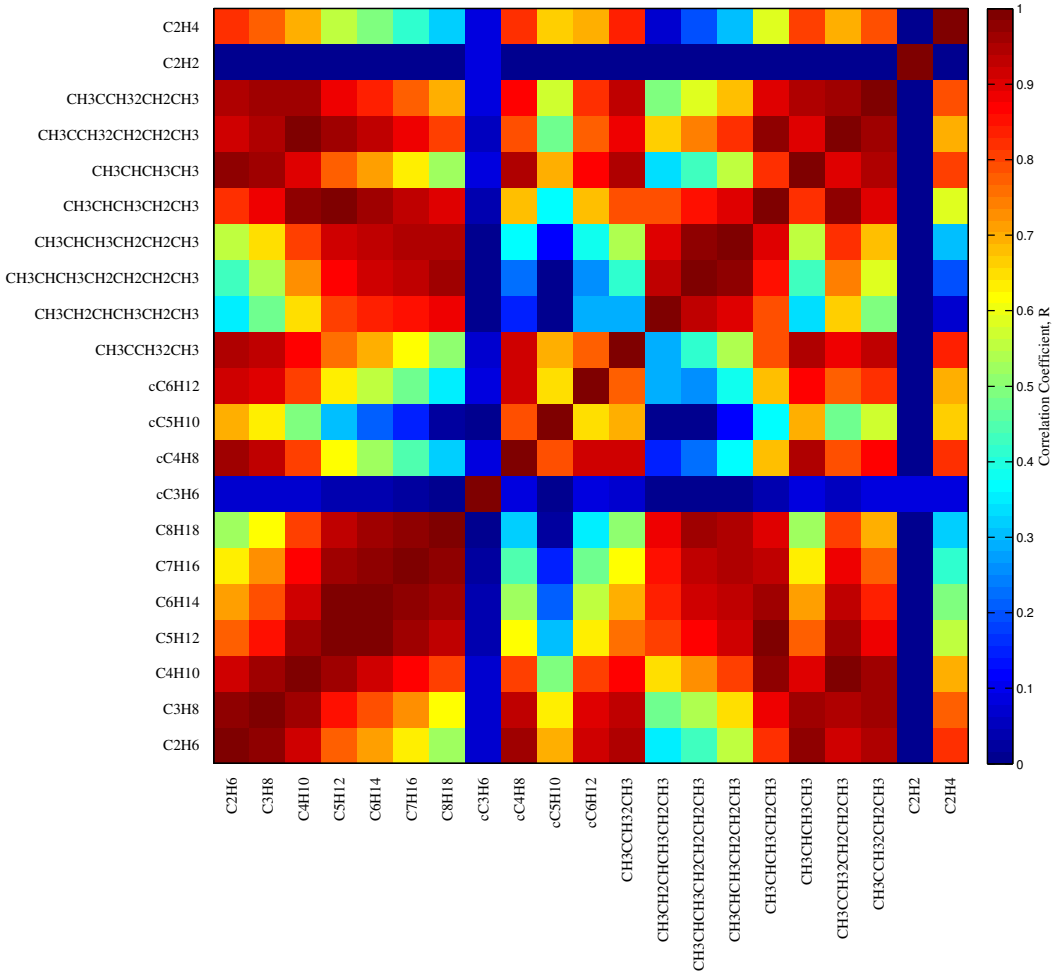

Fig. 19 Correlation coefficient R for the C–C–H bond angles of all molecules examined. The mean bond angle for all C–C–H bond angles was used for those molecules with multiple C–C–H bond angles.

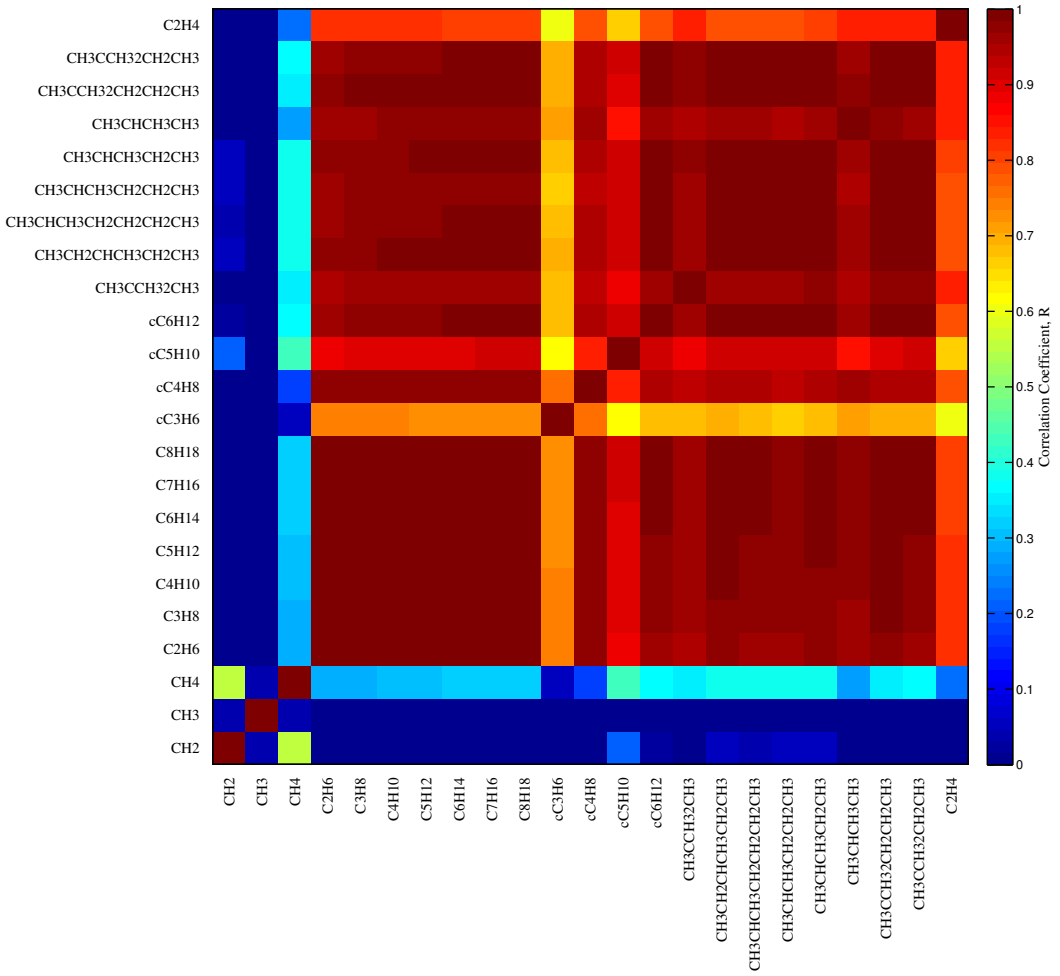

Fig. 20 Correlation coefficient R for the H–C–H bond angles of all molecules examined. The mean bond angle for all H–C–H bond angles was used for those molecules with multiple H–C–H bond angles.

In general, the Latin hypercube sampling provides more expansive coverage of the interatomic potential design space in contrast to the fractional factorial design presented. However, the fractional factorial methods were found to give many of the same trends in terms of parameters. The preceding analysis can be used to better understand the types of molecules or tests that should be included in the training set as well as which ones are duplicates (i.e., correlated). Furthermore, these tests may provide insight as to how to weigh different properties. For instance, Fig. 14 shows that many of the energies of molecules are correlated. This has several implications in terms of weighting the molecule atomization energies. First, if all are equally weighted during multi-objective optimization of the interatomic potential and most molecules are correlated, then clearly the potential will be biased towards capturing the energies of the correlated molecules rather than the noncorrelated molecules. Second, if the weighting is biased towards the correlated/noncorrelated molecules, then the potential may not be as transferable to other molecules that were not weighted as highly during its formulation. Third, if correlations exist, this can be taken advantage of in the training set in terms of reducing the computational cost. That is, the energy of more complex molecule, which might take much longer time to minimize their structure, can be reformulated as functions of less complex molecules and included in the optimization process without running a simulation (i.e., a surrogate model based on a high degree of correlation with less expensive properties). This may have broad implications. A predictive surrogate model for melting temperature based on MEAM parameters and less expensive properties could be used in the optimization process itself, rather than just as a validation test afterwards. These notions and others can be inferred from some of the studies presented within.

4. Summary and Conclusion

An interatomic potential for saturated hydrocarbons[39] using MEAM, a semiempirical many-body potential based on density functional theory, was parameterized to the bond distances, bond angles, and atomization energies at 0 K of a series of alkane structures from methane to n-octane. In this work, the parameters of the MEAM potential were explored through a design of experiments and Latin hypercube sampling approach to better understand how individual MEAM parameters affected several properties of molecules (energy, bond distances, bond angles, and dihedral angles) and the relationship/correlation between various molecules in terms

of these properties. Beyond these results of sensitivity of MEAM parameters and the correlation between different properties, it was also found that reduced-order fractional factorial design of experiment with 64 MEAM combinations yielded similar results to the Latin hypercube sampling design with 1000 MEAM combinations, for instance. The methodology can be easily extended to other potential formulations and can be useful for understanding parameter effects, understanding similarities between properties, and constructing training set tests for the interatomic potential development process.

5. References

1. Daw MS, Baskes MI. Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. Phys Rev Lett. 1983;50(17):1285–1288.

2. Daw MS, Baskes MI. Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B. 1984;29(12):6443.

3. Daw MS, Foiles SM, Baskes MI. The embedded-atom method: a review of theory and applications. Mater Sci Rep. 1993;9(7-8):251–310.

4. Baskes MI. Application of the embedded-atom method to covalent materials: a semiempirical potential for silicon. Phys Rev Lett. 1987;59(23):2666–2669.

5. Baskes MI. Modified embedded-atom potentials for cubic materials and impurities. Phys Rev B. 1992;46(5):2727–2742.

6. Baskes MI, Nelson JS, Wright AF. Semiempirical modified embedded-atom potentials for silicon and germanium. Phys Rev B. 1989;40(9):6085–6100.

7. Lee BJ, Baskes MI. Second nearest-neighbor modified embedded-atom method potential. Phys Rev B. 2000;62(13):8564.

8. Lee BJ, Baskes MI, Kim H, Cho YK. Second nearest-neighbor modified embedded atom method potentials for BCC transition metals. Phys Rev B. 2001;64(18):184102.

9. Lee BJ, Shim JH, Baskes MI. Semiempirical atomic potentials for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, Al, and Pb based on first and second nearest-neighbor modified embedded atom method. Phys Rev B. 2003;68(14):144112.

10. Baskes MI, Srinivasan SG, Valone SM, Hoagland RG. Multistate modified embedded atom method. Phys Rev B. 2007;75(9):094113.

11. Zhang JM, Ma F, Xu KW. Calculation of the surface energy of FCC metals with modified embedded-atom method. Appl Surf Sci2004;229(1):34–42.

12. Zhang JM, Ma F, Xu KW, Xin XT. Anisotropy analysis of the surface energy of diamond cubic crystals. Surf Interf Anal. 2003;35(10):805–809.

13. Baskes MI, Johnson RA. Modified embedded atom potentials for hcp metals. Model Simul Mater Sci Eng. 1994;2:147.

14. Hu W, Zhang B, Huang B, Gao F, Bacon DJ. Analytic modified embedded atom potentials for hcp metals. J Phys: Cond Matt. 2001;13(6):1193.

15. Zhang JM, Ma F, Xu KW. Calculation of the surface energy of BCC metals by using the modified embedded-atom method. Surf Interf Anal. 2003;35(8):662–666.

16. Jelinek B, Groh S, Horstemeyer MF, Houze J, Kim SG, Wagner GJ, Moitra A, Baskes MI. Modified embedded atom method potential for Al, Si, Mg, Cu, and Fe alloys. Phys Rev B. 2012;85(24):245102.

17. Kim HK, Jung WS, Lee BJ. Modified embedded-atom method interatomic potentials for the Fe–Ti–C and Fe–Ti–N ternary systems. Acta Mater. 2009;57(11):3140–3147.

18. Lee BJ, Ko WS, Kim HK, Kim EH. The modified embedded-atom method interatomic potentials and recent progress in atomistic simulations. Calphad. 2010;34(4):510–522.

19. Horstemeyer MF. Integrated computational materials engineering (ICME) for metals: using multiscale modeling to invigorate engineering design with science. Hoboken (NJ): Wiley; 2012.

20. Xiao W, Baskes MI, Cho K. MEAM study of carbon atom interaction with Ni nano particle. Surf Sci. 2009;603(13):1985–1998.

21. Uddin J, Baskes MI, Srinivasan SG, Cundari TR, Wilson AK. Modified embedded atom method study of the mechanical properties of carbon nanotube reinforced nickel composites. Phys Rev B. 2010;81(10):104103.

22. Allinger NL, Yuh YH, Lii JH. Molecular mechanics. The MM3 force field for hydrocarbons. 1. J Amer Chem Soc. 1989;111(23):8551–8566.

23. Lii JH, Allinger NL. Molecular mechanics. The MM3 force field for hydrocarbons. 2. vibrational frequencies and thermodynamics. J Amer Chem Soc. 1989;111(23):8566–8575.

24. Lii JH, Allinger NL. Molecular mechanics. The MM3 force field for hydrocarbons. 3. The van der Waals' potentials and crystal data for aliphatic and aromatic hydrocarbons. J Amer Chem Soc. 1989;111(23):8576–8582.

25. Allinger NL, Chen K, Lii JH. An improved force field (MM4) for saturated hydrocarbons. J Comput Chem. 1996;17(5-6):642–668.

26. Mayo SL, Olafson BD, Goddard WA. DREIDING: A generic force field for molecular simulations. J Phys Chem. 1990;94(26):8897–8909.

27. Brenner DW. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. Phys Rev B. 1990;42(15):9458.

28. Brenner DW, Shenderova OA, Harrison JA, Stuart SJ, Ni B, Sinnott SB. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. J Phys: Cond Matt. 2002;14:783.

29. van Duin ACT, Dasgupta S, Lorant F, Goddard WA. ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A. 2001;105(41):9396–9409.

30. Yu J, Sinnott SB, Phillpot SR. Charge optimized many-body potential for the Si/SiO_2 system. Phys Rev B. 2007;75(8):085311.

31. Shan TR, Devine BD, Hawkins JM, Asthagiri A, Phillpot SR, Sinnott SB. Second-generation charge-optimized many-body potential for Si/SiO_2 and amorphous silica. Phys Rev B. 2010;82(23):235302.

32. Sun H. COMPASS: An ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. J Phys Chem B. 1998;102(38):7338–7364.

33. Somers W, Bogaerts A, van Duin ACT, Huygh S, Bal KM, Neyts EC. Temperature influence on the reactivity of plasma species on a nickel catalyst surface: an atomic scale study. Catal Today. 2013;211(0):131–136.

34. Castro-Marcano F, van Duin ACT. Comparison of thermal and catalytic cracking of 1-heptene from ReaxFF reactive molecular dynamics simulations. Comb Flame. 2013;160(4):766–775.

35. Monti S, Li C, Carravetta V. Dynamics simulation of monolayer and multilayer adsorption of glycine on Cu(110). J Phys Chem C. 2013;117(10):5221–5228.

36. Kim SY, van Duin ACT, Kubicki JD. Molecular dynamics simulations of the interactions between TiO_2 nanoparticles and water with Na+ and Cl-, methanol, and formic acid using a reactive force field. J Mater Res. 2013;28(03):513–520.

37. Liang T, Devine B, Phillpot SR, Sinnott SB. Variable charge reactive potential for hydrocarbons to simulate organic-copper interactions. J Phys Chem A. 2012;116(30):7976–7991.

38. Liang T, Shin YK, Cheng YT, Yilmaz DE, Vishnu KG, Verners O, Zou C, Phillpot SR, Sinnott SB, van Duin ACT. Reactive potentials for advanced atomistic simulations. Ann Rev Mater Res. 2013;43:109–129.

39. Nouranian S, Tschopp MA, Gwaltney SR, Baskes MI, Horstemeyer MF. An interatomic potential for saturated hydrocarbons based on the modified embedded-atom method. Phys Chem Chem Phys. 2014;16:6233–6249.

40. Tschopp MA, Solanki KN, Baskes MI, Gao F, Sun X, Horstemeyer MF. Generalized framework for interatomic potential design: application to Fe–He system. J Nucl Mater. 2012;425(1):22–32.

41. Liyanage LSI, Kim SG, Houze J, Kim S, Tschopp MA, Baskes MI, Horstemeyer MF. Structural, elastic, and thermal properties of cementite (Fe_3C) calculated using a modified embedded atom method. Phys Rev B. 2014;89(9):094102.

42. Valone SM, Baskes MI, Martin RL. Atomistic model of helium bubbles in gallium-stabilized plutonium alloys. Phys Rev B. 2006;73(21):214209.

43. Baskes MI. Atomistic potentials for the molybdenum-silicon system. Mater Sci Eng A. 1999;261(1):165–168.

44. Rose JH, Smith JR, Guinea F, Ferrante J. Universal features of the equation of state of metals. Phys Rev B. 1984;29(6):2963.

45. Foiles SM, Daw MS, Baskes MI. DYNAMO code. Albuquerque (NM): Sandia National Laboratories; 1994.

List of Symbols, Abbreviations, and Acronyms

1NN	first nearest-neighbor
2NN	second nearest-neighbor
ANOVA	analysis of variance
BCC	body-centered cubic
COMB	charge-optimized many-body
COMPASS	condensed-phase optimized molecular potentials for atomistic simulation studies
EAM	embedded-atom method
FCC	face-centered cubic
HCP	hexagonal close-packed
MEAM	modified embedded-atom method
PVT	pressure-volume-temperature
ReaxFF	reactive force field
REBO	reactive empirical bond order
UEOS	universal equation of state

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO L IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	RDRL WMM F M TSCHOPP

INTENTIONALLY LEFT BLANK.